(continued from part 11)

The output characteristics show how the collector current, I_C, varies with the change in the collector emitter voltage, V_{CE}, at different preset base currents.

Once again, it is useful to see how the collector characteristics are derived. *Figure 3b* shows the collector characteristics which are positively biased and are thus shown in the first quadrant of a graph. When the base current is zero, a small collector current will flow, which is known as the common emitter reverse saturation current I_{CEO} (the subscript 'O' stands for open circuit). This is due to the reverse saturation of the collector-base junction. The magnitude of this current will be discussed later in the chapter. The second curve is drawn with a base current of 50 mA.

We have seen that when the transistor is in common emitter mode, there is a current gain of β between the base and the collector. For the transistor under consideration, $\propto$ is 0.975 and β is almost equal to 40. So, when the base current is increased by 50 μA, the collector current will increase by $40 \times 50\ \mu A = 2$ mA. If the base current is taken as constant (e.g. $I_B =$ 50 μA), the curve marked $I_B = 50\ \mu A$ in *figure 3b* shows how the collector current changes with the variation in the collector emitter voltage.

Stepped increases in the base current of 50 μA gives the family of curves shown in *figure 3b*. The emitter output characteristics are similar to those for the common base connection, except that there is now a more marked incline in the current curves. This is because the variation of the collector voltage has an effect which is similar to a very small reduction of the width of the base between emitter and collector and thus, as the collector voltage increases, the current will similarly increase slightly.

Another interesting observation is that the collector current drops to zero when the collector emitter voltage is zero. Since the collector is at the same potential as the emitter, the collector-base junction will be forward biased like the emitter junction. This means that the flow of minority carriers from the base region across the collector junction will stop.

The use of characteristic curves

The choice of transistor for a specific application is determined by operating parameters of the device. Most of this information can be read from the device's **characteristic curves**; the remainder being available in table form. Although information is given for a particular set of operating conditions, it is possible to determine the values of the parameters for *any* operating condition. We shall look at some examples taken from *figures 2* and *3*.

Current gain

Suppose that the current gain of the transistor in *figure 3* is not known, and that you want to find it's value when $V_{CE} = 10$ V and I = 4 mA. Following the $V_{CE} = 10$ V line vertically up the graph of *figure 3b* enables the value of the collector current for each base current variation to be read: with a base current of 50 μA, the collector current is 2 mA; when the base current is 150 μA, the collector current is 6.2 mA. We can therefore see that a variation of 100 μA in the base current brings about a variation of 4.2 mA in the collector current: so the current gain is 42 (4.2 mA $\div$ 100 μA). Comparing the previously defined current gain for this transistor, 40, with this new value shows us that it is correct.

The current transfer values are sometimes shown in transfer characteristic curves. Using the transfer characteristic curve shown in *figure 4a,* where the collector current is a function of the base current, it is possible to deduce the gain at any collector current level. For example, if the current gain is required at a collector current of 6 mA, the corresponding base current is about 150 μA: so the current gain is 40 (6 mA $\div$ 150 μA), which we know to be correct.

The hybrid transfer characteristic is a curve which is used more often. In this case, the collector current is shown as a function of the base emitter voltage (*figure 4b*). This hybrid transfer characteristic curve is particularly useful for applications where the transistor's base voltage is varied to cause a change in the collector current.

Input and output resistances

Transistors are often used to amplify small variable signals which cause a small change in the fixed value of the current that circulates in the transistor. The resistance of this small change will be given by the gradient of the voltage/current graph taken at the static operating point. This is known as the **dynamic resistance**.

The **static resistance**, on the other hand, is represented by the gradient of the line which joins the operating point to the origin. This is important to know when establishing the static conditions of a transistor, and when the transistor is used for switching between different currents.

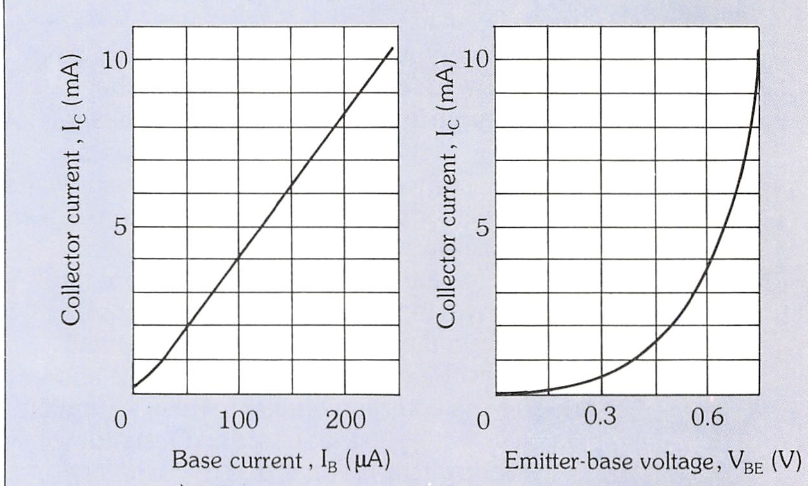
The dynamic resistance that a transistor presents at its input and output must be known in order to determine its appropriateness for an amplifying circuit. These values will be given by the gradients of the input and output characteristics at the static operating point.

For example, let's determine the input and output resistances for the transistor whose characteristics are shown in *figure 2*. The transistor is working with a continuous collector voltage of 10 V and a continuous collector current of 8 mA. First we will have to calculate the input resistance. The gradient of the input characteristic (*figure 2a*) at 8 mA is 0.01 V divided by 4 mA. This gives us an input resistance of 25 Ω.

The common base output resistance is difficult to evaluate accurately from the curves in *figure 2b*, because the gradient is too small. However, for the purpose of this example we can estimate the gradient at 10 V and 8 mA. You will see that a variation of 10 V corresponds to a difference of 0.01 mA. By Ohm's law, this gives us a common base output resistance of 1 MΩ.

Now look at the common emitter characteristics in *figure 3*. You can see that the gradient of the collector characteristic at the desired point is 10 V divided by 0.2 mA. This gives an output resistance of 50 kΩ. A collector current of 4 mA will be produced by a base current of 100 μA. The common emitter input resistance will be equal to the gradient of the input curve

4



at the base current of 100 μA. This is 0.1 V divided by 50 μA which gives 2 kΩ.

These examples show that when a transistor is connected in common base mode, it has an extremely low input resistance (25 Ω) and an extremely high output resistance (1 MΩ). The common emitter arrangement, on the other hand, presents a higher input resistance (2 kΩ) and a lower output resistance (50 kΩ). The transistor configuration that is most suited to a particular use can thus be chosen.

4. Transfer characteristic curves for **(a)** current transfer; **(b)** hybrid transfer.

5. Variations in the reverse saturation currents of low power silicon and germanium transistors.

6. Temperature effects on common-emitter output characteristics: the solid line at 20 °C; the broken line at 100 °C.

The effects of temperature variation

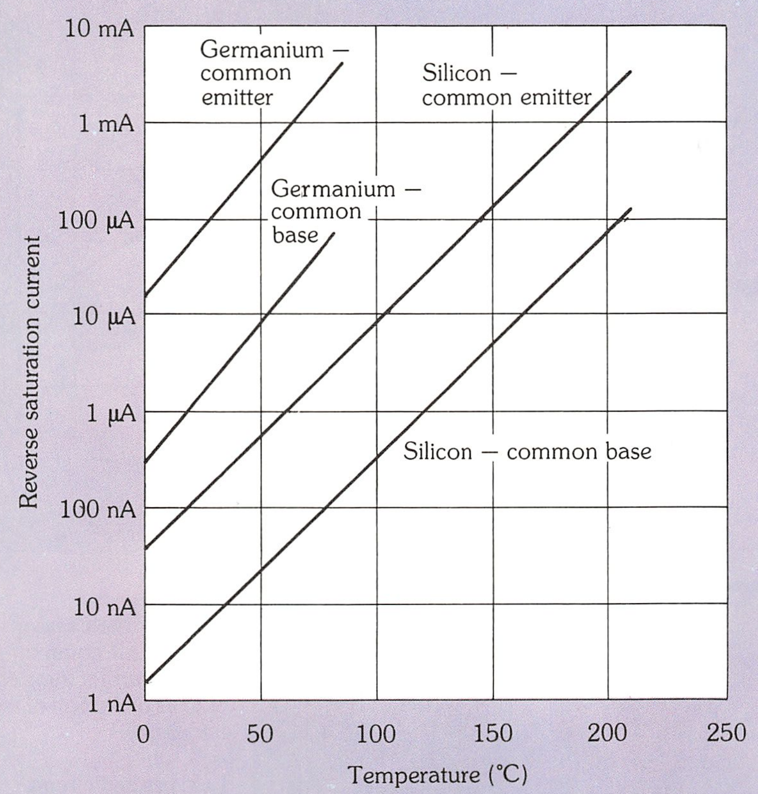
Collector reverse saturation current, I_{CBO}, was mentioned during the discussion of the common base connection. In p-n-p transistors, I_{CBO} flows into the base connection like a negative base current, $-I_{CBO}$. Remember, the + or − sign indicates the direction of current flow.

Common emitter reverse saturation current, I_{CEO}, indicates the leakage current from the collector to the emitter with the base circuit open. A current, $+I_{CBO}$, upon entering the base will neutralize $-I_{CBO}$. The current $+I_{CBO}$ will be amplified by the transistor's current gain, $\boldsymbol{\beta}$, giving a collector current of $\boldsymbol{\beta} \times I_{CBO}$. The common emitter reverse saturation current therefore, is linked to the common base reverse saturation current by the expression:

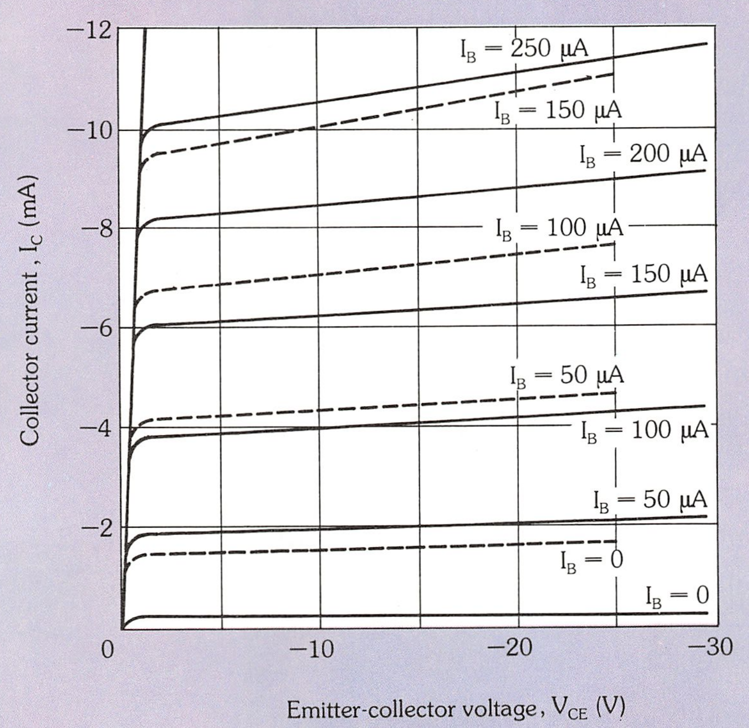
$$I_{CEO} = \boldsymbol{\beta} \times I_{CBO}$$

Since I_{CBO} is very small, I_{CEO} will also be small, but both currents will increase

5



6



exponentially with an increase in temperature. In germanium transistors these currents approximately double for every 9 °C rise in temperature. In silicon transistors they double for each 11 °C increase. *Figure 5* shows the variation in the reverse saturation currents of low power silicon and germanium transistors. Although the common emitter reverse saturation current, I_{CEO}, of a germanium transistor at free air temperature (20 °C) is only approximately 50 μA, there will be an increase of up to 2 mA at 75 °C. In silicon transistors, I_{CEO} at free air temperature is much lower (approximately 100 nA), rising to 2 mA at a temperature of about 200 °C.

This is one reason why silicon transistors are almost always used in preference to germanium transistors; in fact, germanium transistors are now almost obsolete.

Temperature variation also has a considerable effect on a transistor's common emitter current transfer ratio, β. This effect can be seen on a transistor's family of collector characteristics if measurements are taken at different temperatures. The solid curves of *figure 6* indicate typical common emitter output characteristics at a temperature of 20 °C, while the broken line curves show the characteristic at 100 °C. As you can see, the curves are shifted upwards and are further apart.

Frequency characteristics

If an alternating signal is applied to a transistor as input, and the frequency is increased, the output level remains constant at first and then drops. There are two reasons for this.

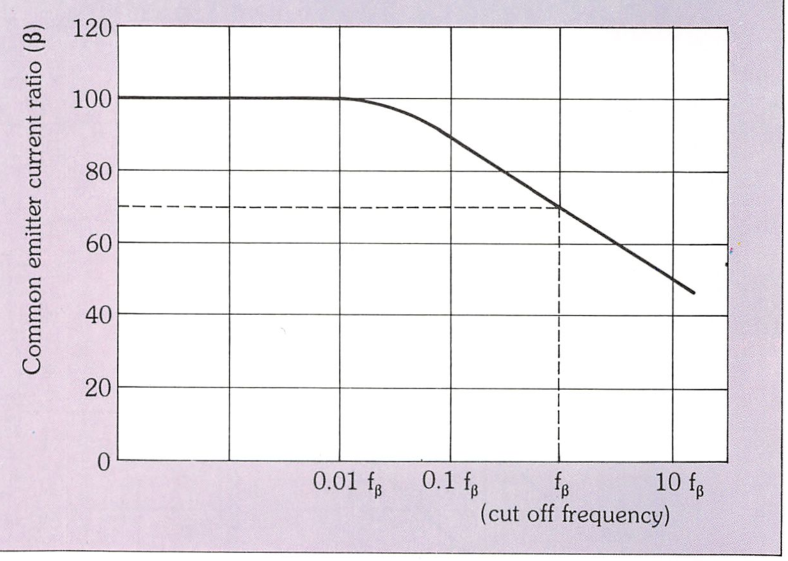
The first is that the current amplification is reduced. After the minority carriers have been emitted, they take a finite time to spread across the base region from the emitter to the collector. This is known as the **transit time** and is proportional to the width of the base region. If the transit time is comparable with the period of the applied alternating signal, the output will be reduced. The typical shape of the frequency characteristic is reproduced in *figure 7* and shows the common emitter current transfer ratio as a function of the frequency. The cut-off frequency, f_β, is indicated, and this is the point at which β drops to about 0.7 times its original low

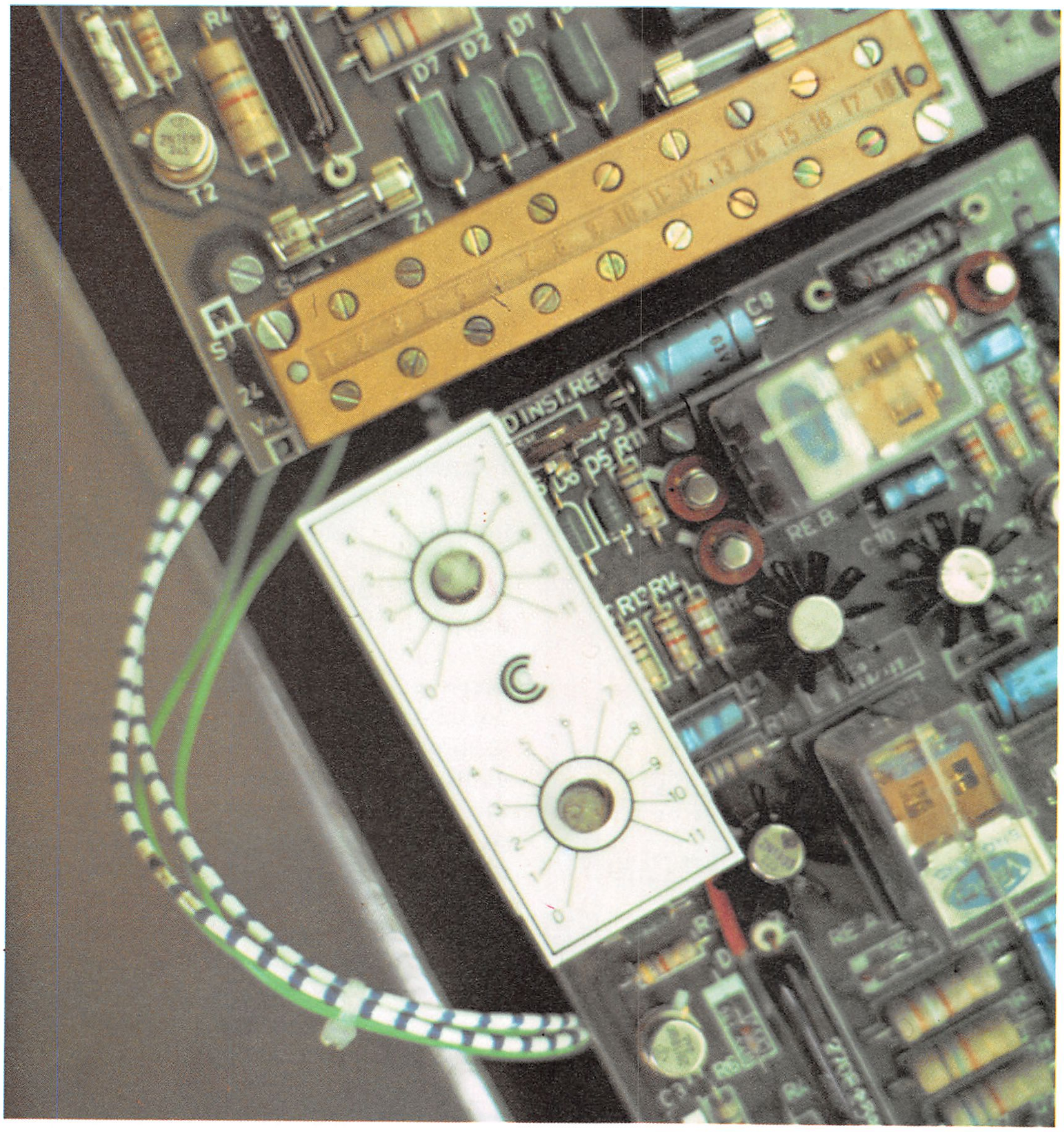
frequency value. Cut-off frequencies of 10 MHz are common in general purpose transistors, but specially manufactured transistors can have cut-off frequencies of up to 2000 MHz. This makes them suitable for use in very high frequency circuits.

The second reason for the decrease in output is that the impedance of the capacitance of the collector junction also diminishes. The effect of the collector capacitance is to short-circuit the collector junction with a low impedance at high frequencies and so decrease the voltage amplification.

As the capacitance is proportional to the area of the junction, this effect can be lessened by reducing the area of the collector junction.

7





Tony Stone Associates

7. Frequency characteristic curve showing the common-emitter current transfer ratio as a function of the frequency.

Left: printed circuit board. Note the two transistors with heat sinks (the black 'frilled' edging).

Maximum ratings of semiconductor devices

Maximum ratings are those limits of a component's operating conditions, beyond which permanent damage or breakdown is caused by operation at too high a voltage or by overheating. The conditions that must be specified for any transistor are the maximum rating of the collector reverse voltage, the reverse voltage of the emitter, the collector current, power handling and the temperature of the collector junction.

The maximum reverse voltage of the collector
In an earlier chapter we saw that if the reverse voltage applied to a p-n junction exceeds certain limits, the phenomenon of avalanche breakdown occurs, i.e. the reverse current will increase rapidly for even a small rise in voltage, generating heat and leading to the possible loss of the device.

The breakdown voltage of a transistor is measured by applying and increasing a reverse voltage until a predetermined rise in current is reached – the rise indicating the beginning of avalanche breakdown. The reverse voltage rating is therefore set *below* this limit, to ensure that the device does not operate in the breakdown region.

Transistors are given a collector breakdown voltage for the common base and common emitter configurations. The common base breakdown voltage, $V_{(BR)CBO}$, is essentially that of the collector junction, and is the breakdown voltage between the collector and the base when the emitter circuit is open.

Since the common emitter reverse saturation current, I_{CEO}, is greater than that for the common base connection, avalanche breakdown occurs at a lower voltage. This means that the common emitter breakdown voltage $V_{(BR)CEO}$ is lower than $V_{(BR)CBO}$. Remember, $V_{(BR)CEO}$ is the voltage between the collector and the emitter, with the base circuit open.

The relationship between the two breakdown voltages depends on the type of transistor under consideration. This is because the breakdown voltage is dependent on the concentration of carriers in the collector and the base, and the amount of current gain a device possesses.

Maximum collector current
Generally speaking, the collector current of a transistor is not limited by its current carriers, but depends on the maximum power level.

Maximum emitter reverse voltage
The emitter junction normally works with a direct bias, but reverse bias of the junction is useful when the collector current is to be blocked. By applying a reverse voltage to the emitter base junction, the collector current can be switched off.

Since the emitter and base regions both have higher concentrations of carriers than the collector region, the reverse breakdown voltage of the emitter junction, $V_{(BR)EBO}$, is usually much lower than that of the collector junction.

Maximum junction temperature and power dissipation
We mentioned earlier that the collector reverse saturation current increases rapidly with temperature. If this temperature increase is excessive, the reverse saturation current becomes so high as to limit the transistor's use. This is because the collector reverse saturation current circulates independently of the device's input conditions.

The junction temperature limit is about 220 °C for silicon transistors.

When a transistor is working, a certain level of power is established in the collector junction which is equal to the product of the voltage across the junction and the current flowing through it. This power causes the generation of heat in the vicinity of the collector junction; the consequent temperature increase depends on the speed at which the heat can be dissipated through the transistor's case to the air.

The heat flow mainly depends on the **thermal resistance** of the materials used, i.e. how much the material resists heat flow. When a transistor is functioning, the collector junction temperature will increase until the heat generated by the electric power is equal to that dispersed by the junction. At this point **thermal equilibrium** is reached. Although there is an amount of power generated in the emitter junction,

this is generally low enough to be disregarded.

Materials with a low thermal resistance are described as good heat conductors. In general, materials which are good heat conductors are also good electrical conductors, and vice versa. This is an important factor in transistor design, as electrical connections can be made to carry current and dissipate heat at the same time.

The overall thermal resistance of a transistor mounted in a circuit is usually indicated as a ratio linking the increase in junction temperature to the surrounding air temperature, per unit of heat dissipated in the transistor. Using this information, and the maximum temperature level, the power level at which the transistor can work at a given temperature can be determined. Subtracting the temperature of the surroundings from the maximum junction temperature gives the permissible temperature increase; dividing this amount by the transistor's thermal resistance gives the permissible power level.

For example, suppose that a silicon transistor has a maximum junction temperature of 200 °C and a thermal resistance of 0-45 °C mW^{-1}. With an environmental temperature of 20 °C, the junction's permissible temperature increase is 180 °C (200 °C – 20 °C). The maximum power dissipation will be 180 divided by 0.45, giving 400 mW. If the environmental temperature was higher, say 100 °C, then the permissible increase would only be 100 °C. The maximum power would then be 100 divided by 0.45, i.e. 220 mW. Component manufacturers sometimes give all this information in the form of a curve, like that shown in *figure 8*.

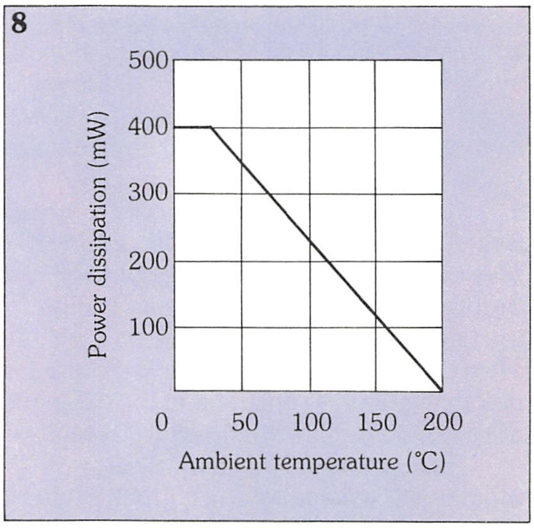
Transistors that are made for high power applications have a very low thermal resistance between the semiconductor element and the surface of the can. The can itself is usually connected to a cooling plate or fin, called a **heat sink**. This means that the surface area in contact with the surrounding air is increased, giving greater heat dissipation and an overall lower thermal resistance.

Similar temperature and power considerations also apply to rectifying diodes. These are reverse biased in the negative half-cycles and forward biased in the positive half-cycles. Most of the power in these devices is produced during the positive half-cycles as a result of the higher levels of current. Because of this, the current limit rating of a rectifying diode is closely linked to the thermal resistance of the device.

Transistor data sheets

To decide which transistor to use for a specific application designers refer to the manufacturer's **data sheets**. Like those already seen for diodes, these data sheets contain the principal characteristics of the device in table form.

The values contained in these tables are calculated under precise operating



8. Relationship between power dissipation and temperature.

conditions, which are always specified. However, there are wide tolerances in these ratings, due to the manufacturing processes used. In spite of this, data sheets are a basic tool for all those involved in electronics, and we shall now look at the information that they supply.

Figures 9 and *10* show the complete data sheet of a transistor in the 2N3250 series. We shall look at the most important information contained on this sheet, which is indicated by the numbers 1 – 7.

1. Current gain

This can be indicated in three different ways: h_{FE}, h_{fe} and β . It is the best known transistor characteristic and indicates the number of times a transistor multiplies the input current to produce the output current.

9. Data sheet for a transistor in the 2N3250 series.

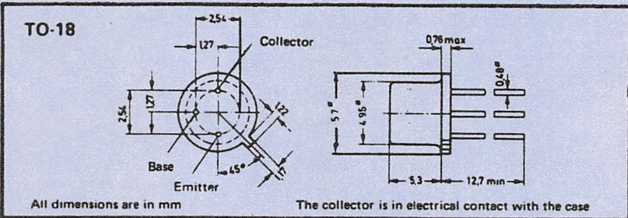
9

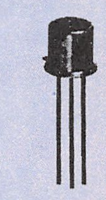
TYPES 2N3250, 2N3250A, 2N3251, 2N3251A
P-N-P SILICON TRANSISTORS

VLB n°511 – December 1973

DESIGNED FOR LOW-POWER SATURATED-SWITCHING AND AMPLIFIER APPLICATIONS

- **Low-Level h_{FE}: 80 Min at 100 μA (2N3251 and 2N3251A)**





***absolute maximum ratings at 25°C free-air temperature (unless otherwise noted)**

	2N3250 2N3251	2N3250A 2N3251A
Collector-Base Voltage	–50 V	–60 V
Collector-Emitter Voltage (See Note 1)	–40 V	–60 V
Emitter-Base Voltage	–5 V	–5 V
Continuous Collector Current	←–200 mA→	
Continuous Device Dissipation at (or below) 25°C Free-Air Temperature (See Note 2)	← 0.36 W →	
Continuous Device Dissipation at (or below) 25°C Case Temperature (See Note 3)	← 1.2 W →	
Storage Temperature Range	–65°C to 200°C	
Lead Temperature 1,6 mm from Case for 60 Seconds	← 300°C →	

***electrical characteristics at 25°C free-air temperature**

	PARAMETER	TEST CONDITIONS		2N3250		2N3250A		2N3251		2N3251A		UNIT
				MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	
$V_{(BR)CBO}$	Collector-Base Breakdown Voltage	$I_C = -10\ \mu A$, $I_E = 0$		–50		–60		–50		–60		V
$V_{(BR)CEO}$	Collector-Emitter Breakdown Voltage	$I_C = -10$ mA, $I_B = 0$,	See Note 4	–40		–60		–40		–60		V
$V_{(BR)EBO}$	Emitter-Base Breakdown Voltage	$I_E = -10\ \mu A$, $I_C = 0$		–5		–5		–5		–5		V
I_{CEV}	Collector Cutoff Current	$V_{CE} = -40$ V, $V_{BE} = 3$ V			–20		–20		–20		–20	nA
I_{BEV}	Base Cutoff Current	$V_{CE} = -40$ V, $V_{BE} = 3$ V			50		50		50		50	nA
h_{FE}	Static Forward Current Transfer Ratio	$V_{CE} = -1$ V, $I_C = -0.1$ mA	See Note 4	40		40		80		80		
		$V_{CE} = -1$ V, $I_C = -1$ mA		45		45		90		90		
		$V_{CE} = -1$ V, $I_C = -10$ mA		50	150	50	150	100	300	100	300	
		$V_{CE} = -1$ V, $I_C = -50$ mA		15		15		30		30		
V_{BE}	Base-Emitter Voltage	$I_B = -1$ mA, $I_C = -10$ mA	See Note 4	–0.6	–0.9	–0.6	–0.9	–0.6	–0.9	–0.6	–0.9	V
		$I_B = -5$ mA, $I_C = -50$ mA			–1.2		–1.2		–1.2		–1.2	V
$V_{CE(sat)}$	Collector-Emitter Saturation Voltage	$I_B = -1$ mA, $I_C = -10$ mA			–0.25		–0.25		–0.25		–0.25	V
		$I_B = -5$ mA, $I_C = -50$ mA			–0.5		–0.5		–0.5		–0.5	V
h_{ie}	Small-Signal Common-Emitter Input Impedance	$V_{CE} = -10$ V, $I_C = -1$ mA, $f = 1$ kHz		1	6	1	6	2	12	2	12	kΩ
h_{fe}	Small-Signal Common-Emitter Forward Current Transfer Ratio			50	200	50	200	100	400	100	400	
h_{re}	Small-Signal Common-Emitter Reverse Voltage Transfer Ratio				10×10^{-4}		10×10^{-4}		20×10^{-4}		20×10^{-4}	
h_{oe}	Small-Signal Common-Emitter Output Admittance			4	40	4	40	10	60	10	60	μmho

NOTES: 1. These values apply between 0 and 200 mA collector current when the base-emitter diode is open-circuited.
2. Derate linearly to 200°C free-air temperature at the rate of 2.06 mW/deg.
3. Derate linearly to 200°C case temperature at the rate of 6.9 mW/deg.
4. These parameters must be measured using pulse techniques. $t_p = 300\ \mu s$, duty cycle $\leq 2\%$.

*Indicates JEDEC registered data

As an example, suppose that a transistor is connected in common emitter mode, and that an input to the base of 1 mA produces an output of 100 mA from the collector. The current gain in this case is 100. The difference between h_{FE} and h_{fe} is that h_{FE} indicates the static amplification for direct current, and h_{fe} shows the current gain for small AC signals: both of these are given in the data sheet. These factors are extremely variable, even for the same transistor, and must be given specifying V_{CE}, I_C and the working frequency.

2. Noise figure (NF)

Noise in an electrical circuit is an unwanted signal. The background hisses and crackles on radios are audible noise signals. Every electronic device produces noise to some extent, and it is generally caused by a turbulent and irregular flow of electrons.

A transistor's output current is slightly

10

10. Second half of data sheet shown in figure 9.

TYPES 2N3250, 2N3250A, 2N3251, 2N3251A P-N-P SILICON TRANSISTORS

***electrical characteristics at 25°C free-air temperature (continued)**

PARAMETER		TEST CONDITIONS	2N3250 2N3250A MIN	2N3250 2N3250A MAX	2N3251 2N3251A MIN	2N3251 2N3251A MAX	UNIT
$\lvert h_{fe} \rvert$	Small-Signal Common-Emitter Forward Current Transfer Ratio	$V_{CE} = -20$ V, $I_C = -10$ mA, f = 100 MHz	2.5		3		
f_T	Transition Frequency	$V_{CE} = -20$ V, $I_C = -10$ mA, See Note 5	250		300		MHz
C_{obo}	Common-Base Open-Circuit Output Capacitance	$V_{CB} = -10$ V, $I_E = 0$, f = 100 kHz		6		6	pF
C_{ibo}	Common-Base Open-Circuit Input Capacitance	$V_{EB} = -1$ V, $I_C = 0$, f = 100 kHz		8		8	pF
$r_b'C_c$	Collector-Base Time Constant	$V_{CE} = -20$ V, $I_C = -10$ mA, f = 31.8 MHz		250		250	ps

NOTE 5: To obtain f_T, the $\lvert h_{fe} \rvert$ response with frequency is extrapolated at the rate of −6 dB per octave from f = 100 MHz to the frequency at which $\lvert h_{fe} \rvert = 1$.

***operating characteristics at 25°C free-air temperature**

PARAMETER		TEST CONDITIONS	2N3250 2N3250A MAX	2N3251 2N3251A MAX	UNIT
NF	Spot Noise Figure	$V_{CE} = -5$ V, $I_C = -100$ μA, $R_G = 1$ kΩ, f = 100 Hz	6	6	dB

***switching characteristics at 25°C free-air temperature**

PARAMETER		TEST CONDITIONS†	2N3250 2N3250A MAX	2N3251 2N3251A MAX	UNIT
t_d	Delay Time	$I_C = -10$ mA, $I_{B(1)} = -1$ mA, $V_{BE(off)} = 0.5$ V,	35	35	ns
t_r	Rise Time	$R_L = 275$ Ω, See Figure 1	35	35	ns
t_s	Storage Time	$I_C = -10$ mA, $I_{B(1)} = -1$ mA, $I_{B(2)} = 1$ mA,	175	200	ns
t_f	Fall Time	$R_L = 275$ Ω, See Figure 2	50	50	ns

†Voltage and current values shown are nominal; exact values vary slightly with transistor parameters. Nominal base current for delay and rise times is calculated using the minimum value of V_{BE}. Nominal base currents for storage and fall times are calculated using the maximum value of V_{BE}.

*PARAMETER MEASUREMENT INFORMATION

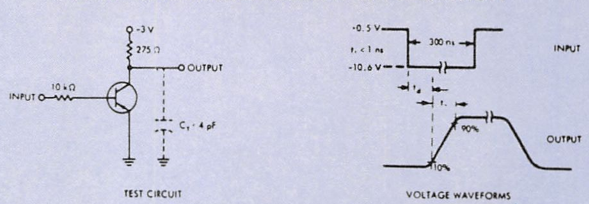


FIGURE 1—DELAY AND RISE TIMES

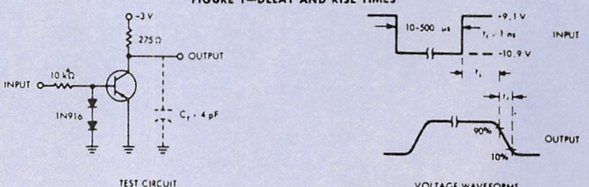


FIGURE 2 — STORAGE AND FALL TIMES

NOTES: a. The input waveforms are supplied by a generator with the following characteristics: $Z_{out} = 50$ Ω, duty cycle = 2%.

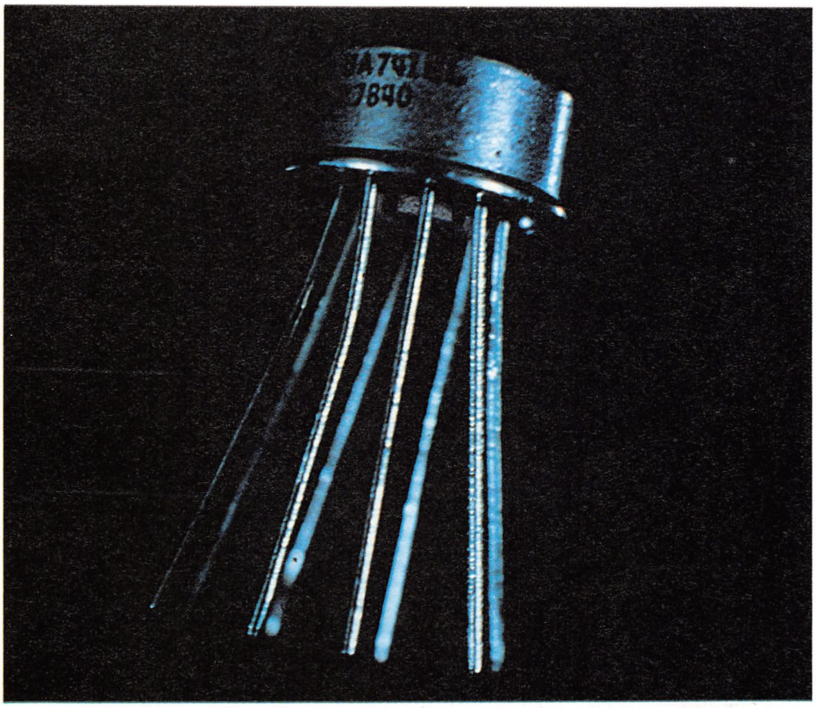
b. Waveforms are monitored on an oscilloscope with the following characteristics: $t_r \leq 1$ ns, $R_{in} \geq 100$ kΩ.

*Indicates JEDEC registered data

irregular, and is not a perfect replica of the input current. The noise figure of a transistor gives an idea of how irregular and imperfect the working current is, and is measured in relative units of intensity known as **decibels** (dB). A decibel is a logarithm of the ratio of the output level to the input level. A level of N dB is defined as:

$$N\,dB = 10 \log_{10} (P_2/P_1)$$

where P_2 is the output power level and P_1 is the input power level.



Above: a differential amplifier.

3. Maximum power dissipation
This is the power wasted by conversion into heat, as we have seen (measured in watts). Circuit designers need to know at what temperature a transistor will malfunction or burn out. The maximum bearable level of collector current (I_c) is also specified, as exceeding this value will generate a destructively high temperature.

4. Conductance
Conductance is an indication of how easily current flows in the transistor when switched full on (i.e. saturated or bottomed).

A measure of a transistor's input conductance is given by the base-emitter voltage V_{BE}, required to keep a specified collector current flowing.

Similarly, a small collector-emitter voltage is needed to keep the collector current flowing when the transistor is saturated. This voltage is known as the **collector-emitter saturation voltage**, $V_{CE(sat)}$.

5. Leakage current
This is an unwanted current which flows through the transistor when it should be non-conducting. The leakage currents for a transistor have to be known for the three different modes of connection – common base, common collector and common emitter. Leakage currents are measured for data sheets with one terminal connected to a specific voltage, and are known as the the **cut-off currents**. The cut-off currents I_{CEV} and I_{BEV} are usually given. The 'V' stands for the constant voltage applied to the third terminal.

6. Breakdown voltage
The breakdown voltage is the highest voltage a device can handle when it is switched **off**. If this limit is exceeded, the device breaks down and is destroyed. The voltage $V_{(BR)CBO}$, for example, indicates the breakdown voltage of the collector base junction when the emitter is open. The data sheet in *figure 9* also shows the collector emitter and the emitter base breakdown voltages.

7. Operating speed
This indicates the time taken for the transistor to switch on and off. This characteristic is specified with respect to the emitter collector working circuit, as it refers to the speed at which the working circuit responds to the variations in the control current. In the case of switching transistors, this is the time needed to switch the device on (usually from 10% up to 90% of the full voltage) or to switch it off (usually from 90% down to 10% of the full voltage), referred to as the **rise time** and the **fall time** respectively.

Amplifying transistors, on the other hand, are given a **frequency** specification, rather than a switching specification. The most common characteristic being the **transition frequency (f_T)**. This is the highest frequency at which the transistor can usefully operate as an amplifier. Occasionally, neither switching times nor frequencies are given, however other specifications will indicate how the device will operate at frequencies defined in the testing conditions.

There are a large number of standard semiconductor specifications (as well as many others which are non-standard). However, with the important characteristics covered in this chapter, you should be well equipped to understand any transistor's specification and work out how it should perform in a circuit under any given set of operating conditions.

Glossary

breakdown voltage	maximum reverse voltage that a transistor can tolerate, before avalanche breakdown occurs
conductance	an indication of how easily current flows in a device when it is switched on
reverse saturation current I_{CEO}	indicates the leakage current from the collector to the emitter with the base open in a common emitter circuit
dynamic resistance	the resistance of a device under operating conditions
operating speed	in switching transistors the time taken to switch on and off. Sometimes known as the rise time and the fall time
gain	in a transistor current gain is indicated by h_{FE}, h_{fe} or β . This indicates the number of times a transistor multiples the input current to produce the output current
heat sink	device used to dissipate unwanted heat, especially from power transistors
leakage current	an unwanted current that flows through a transistor
maximum junction temperature	maximum temperature of the collector junction, which must not be exceeded, as the reverse saturation current would become so high as to limit the use of the transistor
maximum power dissipation	indicates the amount of heat that a transistor can dissipate from its case to the outside air, before a point is reached where the device would malfunction or burn out
thermal resistance	a measure of the degree to which a material can resist heat flow. If a material has low thermal resistance it is a good heat conductor
transit time	the time taken for minority carriers to spread across the base region from the emitter to the collector – proportional to the width of the base region
transition frequency (f_τ)	the highest frequency at which a transistor can operate as an amplifier

7

BASIC COMPUTER SCIENCE

Data files

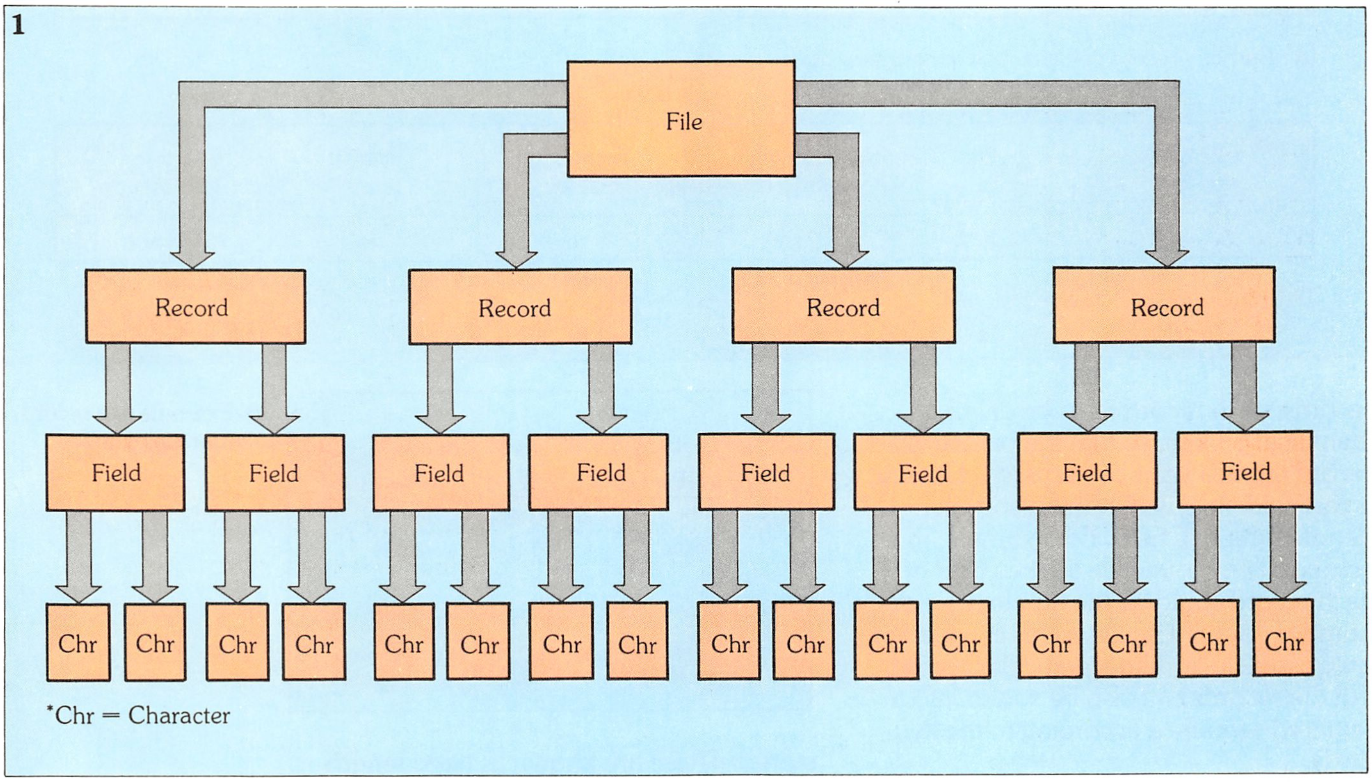
Mass memory storage

In *Computer Science 6* we looked closely at how data is structured and represented within the central memory of a computer. We can continue the discussion by studying the different ways in which data structures are stored in **auxiliary** or **mass memory** media, such as magnetic tape and disk and magnetic drum. This type of physical memory medium is usually remote from the central processor memory.

As in central memory, the physical organization of data within auxiliary memory has no immediate correspondence with the way data is used by the computer. Between the physical representations of data in auxiliary memory and the user is an interface consisting of the **programming language** (e.g. BASIC, FORTRAN, Pascal) and the computer's **operating system.** The operating system is a set of programs which control the computer and its resources.

The higher the level of language and the more powerful the operating system, the more simple and flexible it is for a computer user to store and retrieve data. Often, with complex computers, the user may be totally ignorant regarding the exact whereabouts of the data.

1. A hierarchical family tree of a file.



Organisation of data

Sets of data stored in mass memory are called **files.** A single reel of magnetic tape could hold one or more computer files containing information that needs to be kept for a long time, for example, a list of company employees or sales figures for the tax year. These files are logically sequ-

enced data structures and can be considered either as tables or linear lists where each element is called a **record**.

Each record may include many data items, called **fields,** which are not necessarily related to each other. Each field can be further subdivided into individual **characters.** A hierarchical family tree of a file can be drawn and one is shown in *figure 1*.

When a file is structured as a table of records, each record must be given an identification **key,** so that the particular record required can be accessed. Usually, a particular field within the record is denoted the **key field.** For example, in a company employee file the key field of each record could be the employee's clock number. Identifying key fields in a file not only assists in finding a particular record, but the records can also be sorted into a required sequence according to the key fields.

Alternatively, a linear list of records which are in a sequence can be scanned sequentially until a desired record is found. *Figure 2* shows two records which are divided into fields; maximum character lengths of each field are indicated. In the examples shown, the upper record (*figure 2a*) is that of a company employee; the lower record (*figure 2b*) contains information relating to a customer. An example of a data record comprising seven fields is shown in *figure 3* (the maximum number of characters allowed in each field is also shown). A graphical representation of a field is shown in *figure 4* – this illustrates how physical records are combined to form a **logical record.** As you can see, each field holds data relating to one particular aspect of the record.

In the examples shown, each field (and therefore each record) is of a fixed maximum size, which automatically limits the amount of information that can be stored. These are known as **fixed length records.** However, in the majority of applications, particularly commercial, **variable length records** prove more useful. Considerable storage space can be wasted in a fixed length record system if every field is not fully utilised, so although variable length records pose more problems for the user they are often preferred. (The first field within each variable record is often used to indicate the record's length.)

2. Examples of two records divided into fields.

2

Works number	Name		Address		Personal information	
	Surname	First names	Street	City/town	Place of birth	Date of birth
8 chrs	30 chrs		40 chrs		23 chrs	
	10 chrs	20 chrs	25 chrs	15 chrs	15 chrs	8 chrs

(a) Part of a record containing an employee's personal data

Client code	Article code	Quantity ordered	Salesman's code	Payment terms	Other information
6 chrs	8 chrs	6 chrs	5 chrs	4 chrs	6 chrs

(b) Record of sales to a customer

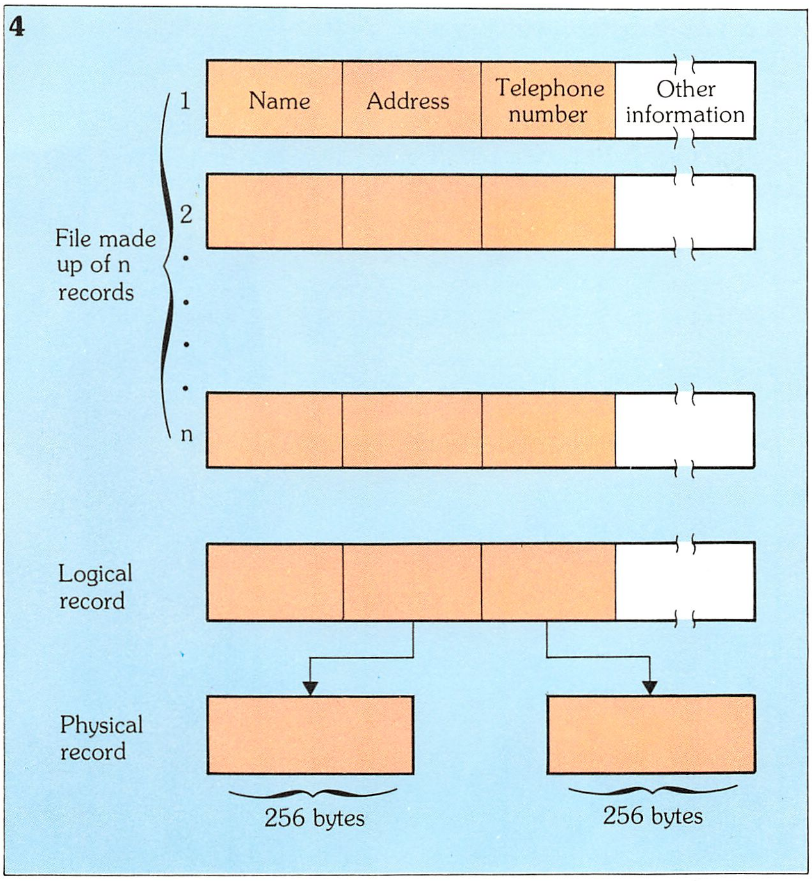
3. Example of a record comprising 7 fields.

3

Works number	8 chrs	numbers
Surname	10 chrs	letters
First name	20 chrs	letters
Street	25 chrs	letters
City/town	15 chrs	letters
Place of birth	15 chrs	letters
Date of birth	8 chrs	numbers

Structuring data

Auxiliary memory handles data in block



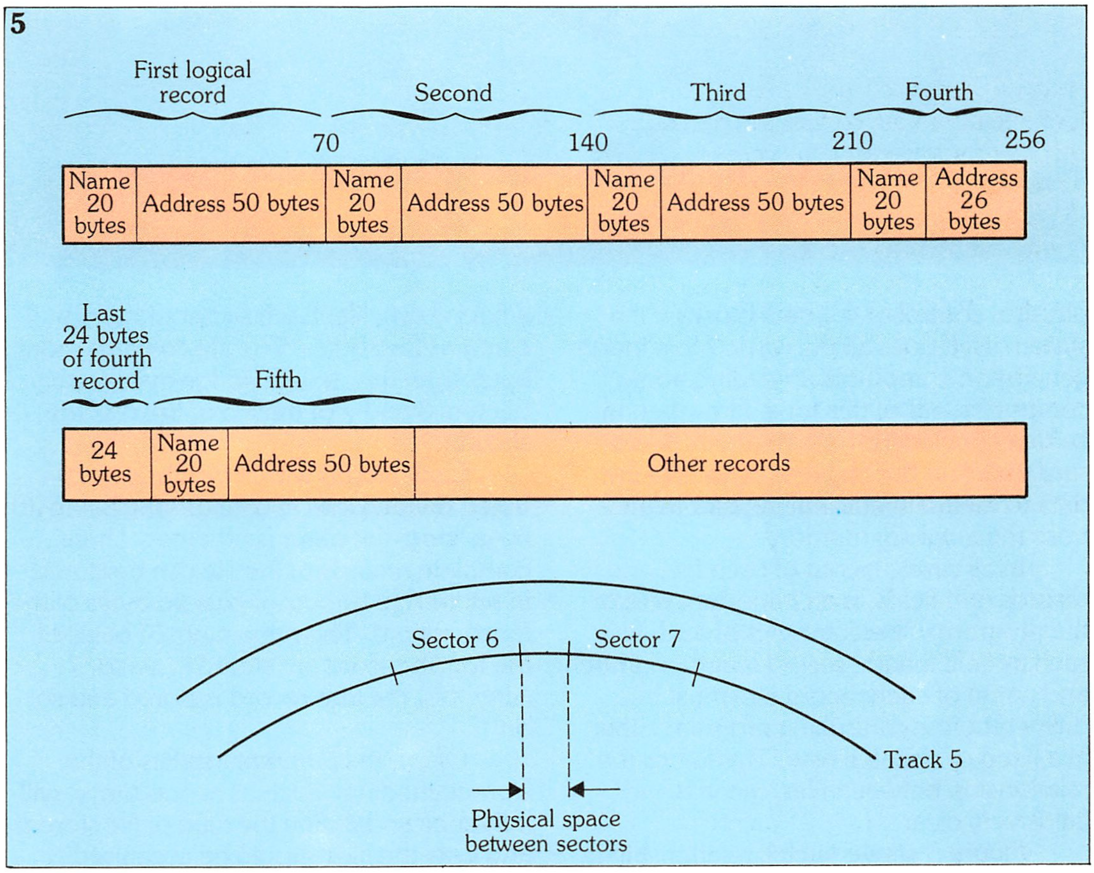
form whereas the computer handles and stores data in its central memory in bytes. The length of these physical blocks varies according to the type of auxiliary memory used. This presents a problem – how to interface the byte-by-byte data structures of the computer to the block-by-block storage methods of the different auxiliary memories.

Physical transfer of data between the computer and the auxiliary memory generally occurs in a byte-by-byte manner, each byte being temporarily stored in the **buffer registers** – these are byte-length stores, one in the computer, one in the auxiliary memory. Transfer of bytes can be either serial (i.e one bit at a time), or parallel (i.e. many bits at once) depending on the design of the hardware. Both buffers are controlled by the computer software.

The way data is written into, or read from, each buffer is dependent upon the programming language and operating system used. For example, a high level language used in conjunction with a powerful operating system ensures that the user need not be concerned with the

4. Graphical representation of a field illustrating how physical records are combined to form a logical record. Each field holds data relating to one particular aspect of the record.

5. Logical records on a floppy disk.



physical control of the data blocks – the system itself does all the work. Most fourth generation computers, including home computers, are of this type. However, in certain computers it is sometimes necessary to define the physical aspects of the data to maintain efficient transfer to and from the auxiliary memory.

In all cases, layout of each file, and its records and fields, is an important part of the program. Characteristics of each file, such as field length, record length, number and name of each record etc. must be defined by the controlling program either in a fixed or variable way. This keeps the relationship between files, records and fields very clear.

Figure 5 shows such a relationship in a floppy disk file. Each sector of the disk contains 256 bytes. The file contains more bytes than this, so will bridge over sectors. Sectors 6 and 7 of track 5 of the disk are shown.

Each record within the file is of 70 bytes, divided into two fields: one being 20 bytes long; the other is 50 bytes. Three complete records of the file can be stored in sector 6 (210 bytes), plus 46 bytes of the fourth record. The remaining 24 bytes of the fourth record are stored in sector 7. After this, the fifth record is stored and so on.

Often, the *inherent* aspects of the storage medium, such as access times, will influence and define the *type* of file stored, and how the files should be organised.

Above: the computer as a professional tool.
(Photo: Honeywell).

Files on magnetic tape

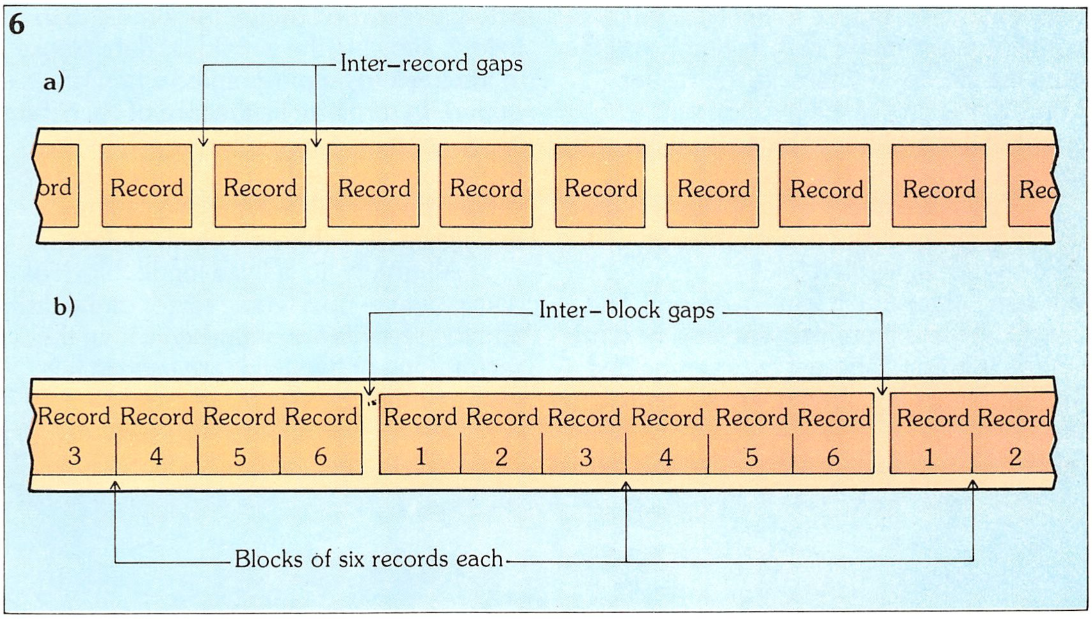
Magnetic tapes can only record data files serially and are therefore known as **serial storage media.** Because of this they can be seen as linear lists and can be processed and stored as such.

Records are **written,** i.e. recorded onto the tape, by the computer. One method separates each record with a **gap** in which nothing is recorded. This **inter-record gap** (IRG) allows the tape mechanism to slow down and stop after one record, and then start ready for the next (see *figure 6a*).

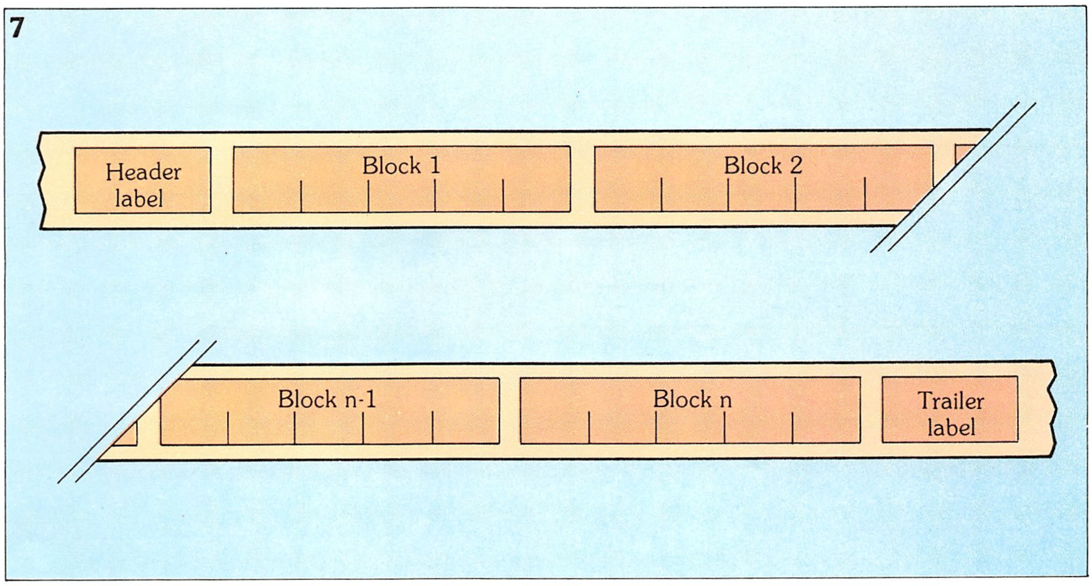
Obviously, the use of IRGs wastes a great deal of tape. Another writing method, however, groups records together into **blocks.** Each block of records must then have an **inter-block gap** (IBG) to allow for the tape stopping and starting (see *figure 6b*). This method reduces the amount of wasted tape and also the data transfer time.

File organisation on tape

Records can be written onto tape to form files in two distinct ways: a **serial** file in which records are written without any relationship existing between them; or a **sequential** file where records are written in a set sequence, according to key fields.



6. A collection of records on magnetic tape showing **(a)** inter-record gaps; **(b)** inter-block gaps.



7. Format of a single volume file with header and trailer labels.

In both cases, however, the moment when the physical block of data is written onto tape does not interest the computer user. The user controls the records within the program, inserting and removing data while the record is held in the computer. But when the program is run, all records are written onto tape.

A tape file can occupy more than one reel of tape in which case it is called a **multi-volume file.** Alternatively, a single tape might contain several files. In either situation, each file needs indentifying (i.e. file volume, name, beginning of records, end etc.) by means of a **label.** *Figure 7* illustrates the possible format of a single-volume file with labels. The **header label**, which identifies the file to the computer, contains information such as name and the date the file was written etc; the **trailer label,** at the end of the file, contains a count of the number of records on file and the reel number where necessary.

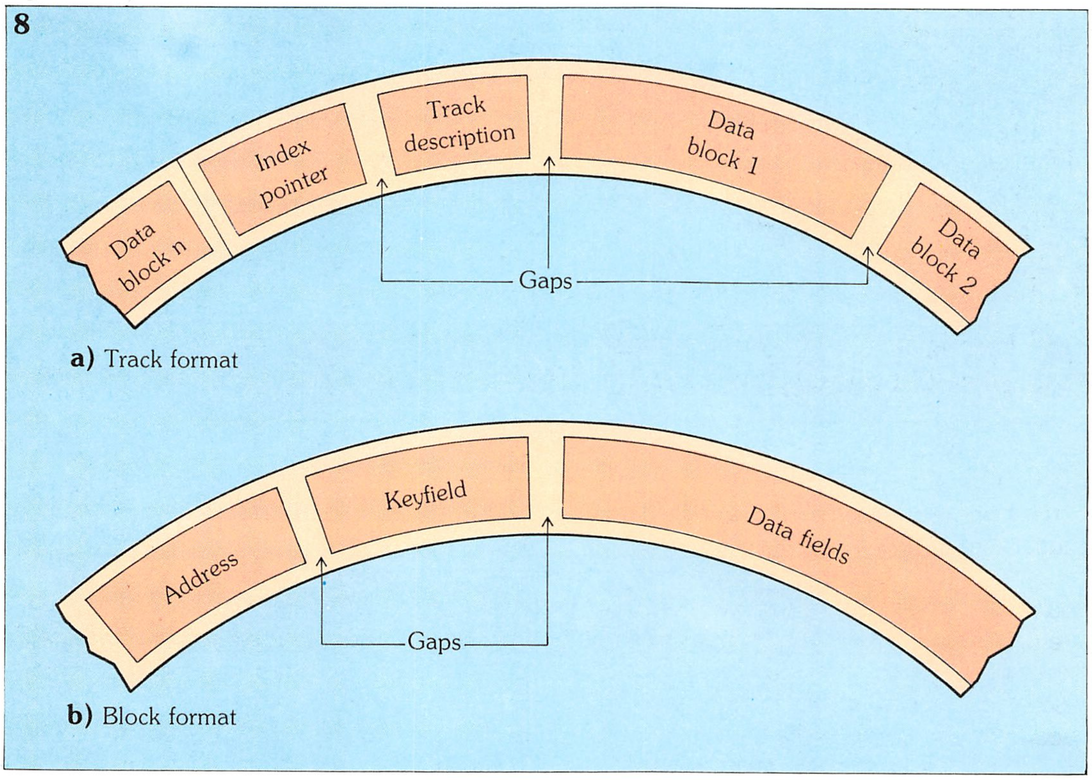
Updating magnetic tape files is quite a complex procedure because the updated data cannot be written back onto tape in the same place that it was read from. The complete file is therefore read and held in central memory where it can be modified. The updated file is then rewritten onto tape.

Magnetic disk files

Disk data storage provides a good example of a **direct access** (sometimes known as **random access**) medium because files can be accessed in a non-serial manner. So, it doesn't matter in what order the files are written onto the disk, access is made rapidly. If you compare this to the serial access of magnetic tape then the advantage in terms of reduced access time becomes clear.

If records are to be recorded as a variable length block (similar to those used with magnetic tape) an **index pointer** is recorded which indicates the physical start of the track. A description of what is actually recorded on the track must then follow, allowing the individual data blocks in the track to be pin-pointed when required. In turn, the first record of each data block contains information which indexes the separate records in the block. *Figure 8a* shows the possible format of a disk track and *figure 8b* is that of a block of data.

Alternatively, a fixed length block of data can take up a whole sector of the disk and the records are organised within the sector. One or two fields are used in this case to index the sector.



8. Magnetic disk storage showing **(a)** track format; **(b)** block format.

Serial disk organisation

There are four basic ways in which files on disk can be organised. The **serial** method is similar to that used with magnetic tape; where records within each file are written one after the other. The records are generally **unsorted,** i.e. in no specific order. When accessing serial files it is necessary to read the whole file, record by record, until the required one is found.

Above: magnetic disk units from a mini-computer. (Photo Honeywell).

Modifications can be made to the data on the disk and information in fields can be updated or erased. However, it must be remembered that each field, once recorded, has a fixed number of character spaces and modifications must not increase the field length. Likewise, erasure of a field must be indicated by a specific field. It is also sometimes possible to add new records at the end of a file.

A complete block of logical records and, in fact, a complete file can be modified and rewritten onto the disk in the same place. But the computer software must have knowledge of the exact position on the disk and length of data to be rewritten.

It is obvious that serial file organisation does not make particularly efficient use of either computer time or disk space.

Sequential disk file

The **sequential** method of disk storage allows records to be written as a defined sequence. However, the records do not need to be stored in physical sequence, because the first field of every record contains not only a key to what information is held in the record, but also a **pointer** which indicates which record is next. Such a method of data storage is known as a **linked list** of data items and a possible

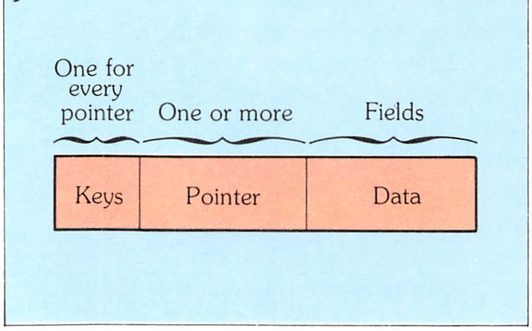
format is shown in *figure 9*.

Existence of this pointer means that modifications can be easily made to the file because, as *figure 10* shows, it is a simple matter to alter the address indicated by the pointer to the modified or new address. *Figure 10a* shows how a record can be added to a sequential file and *figure 10b* shows how a record can be deleted. However, a new record does require storage area on the disk so the disk must not be full for this sort of modification to occur. Similarly, record deletion leaves an unused space on the disk, which can be re-used to store a new record.

To overcome these problems, some computer systems automatically create a special file on the disk known as a **space list** which indicates the empty or unused spaces on the disk. This list itself is controlled in a FILO (first in, last out) structure so that whenever a record is to be added to a file, the space it is to be recorded in is taken from the top of the list. The list is then updated so that the next available space becomes top of the list, as shown in *figure 11* where the record with the key B is to be added to the data block, between records A and C. First, the space list indicator is changed to point to the second free record in the space list, S2. Then the pointer of record A in the data block is changed to indicate S1 (i.e. record B). Finally, the pointer of S1 (or B) is changed to indicate record C.

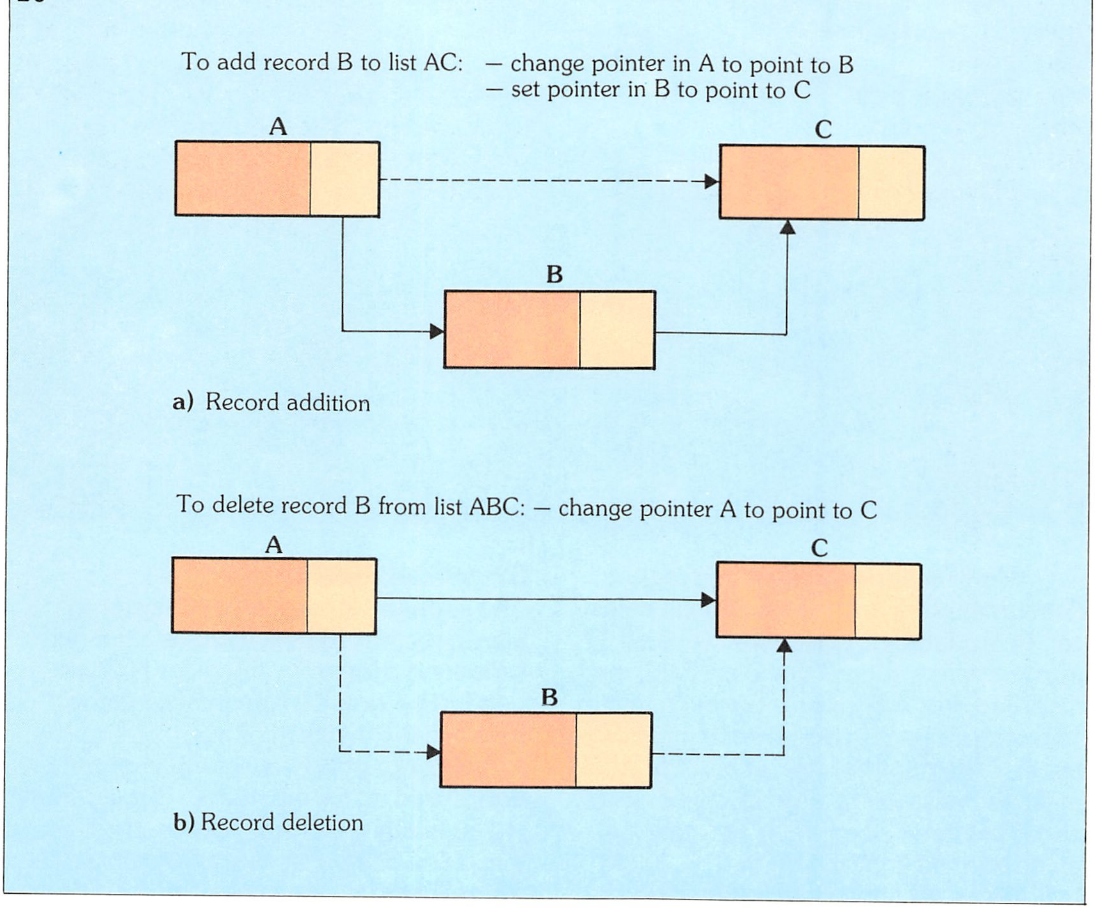
When a record is to be deleted from a file it is automatically returned to the top of the space list. Thus, records at the top of

9



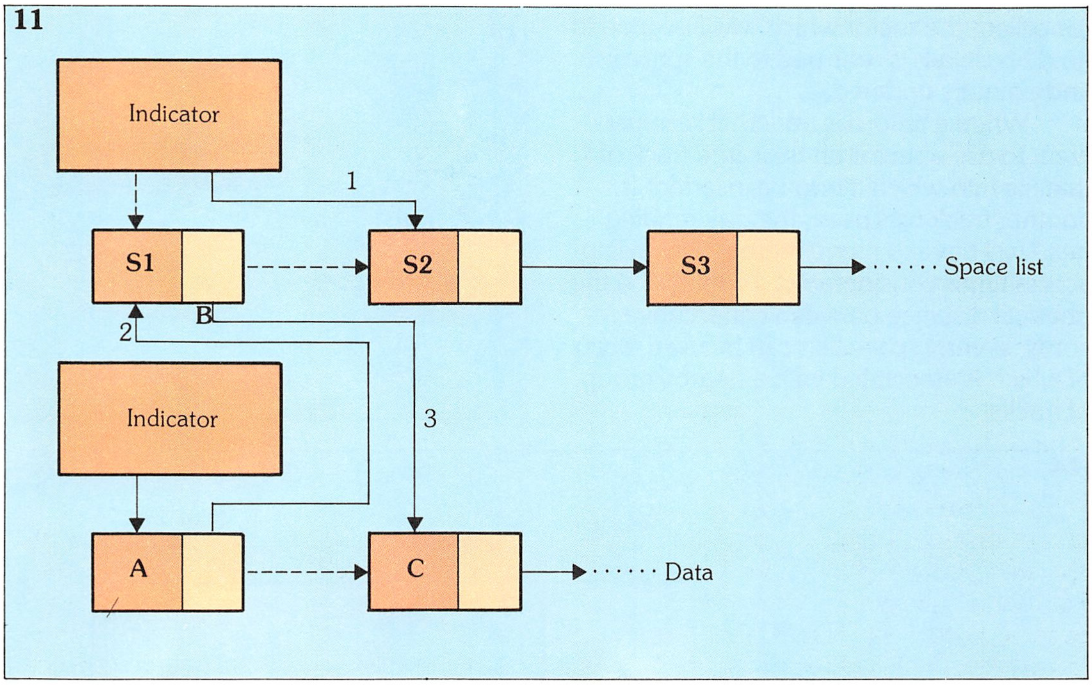
9. Possible format for a linked list of data items.

10



10. (a) Record addition and (b) record deletion using the pointer.

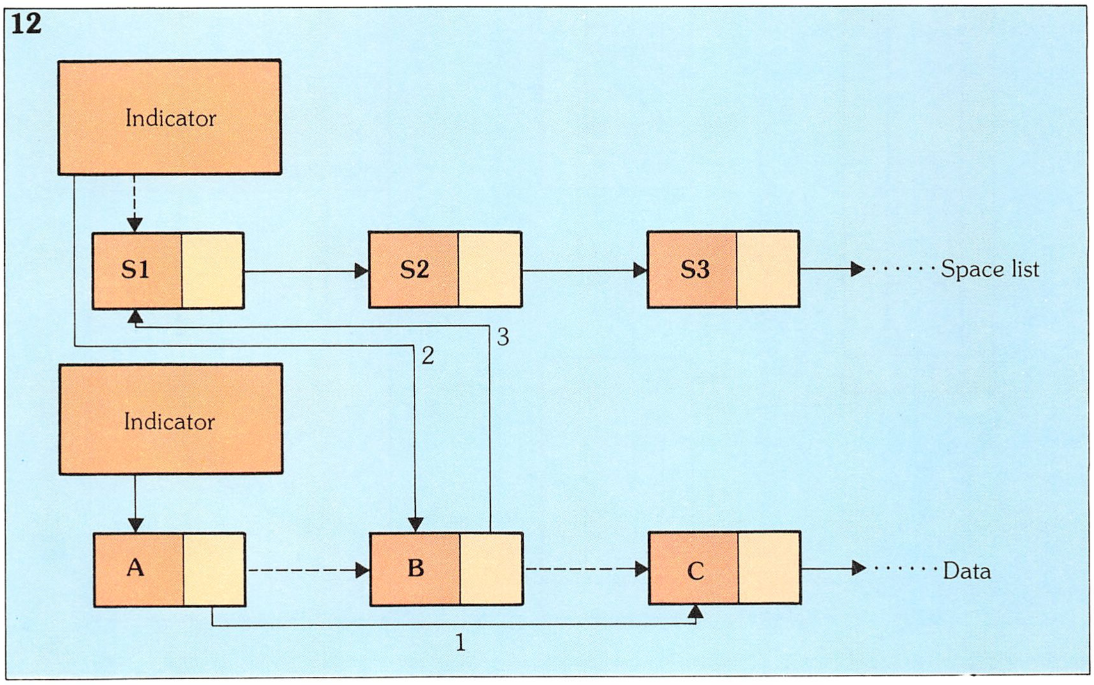
11. The space list is a special file on disk which indicates the empty or unused spaces, thereby saving disk space.



the list will be more likely to be used than those at the bottom. This is shown in *figure 12*, where record B is to be cancelled from the data block. First, the pointer of record A is changed to indicate C. Next, the indicator of the space list is changed to point to the record to be deleted (i.e. B), and finally the pointer of record B is changed to point to the top of the space list, S1.

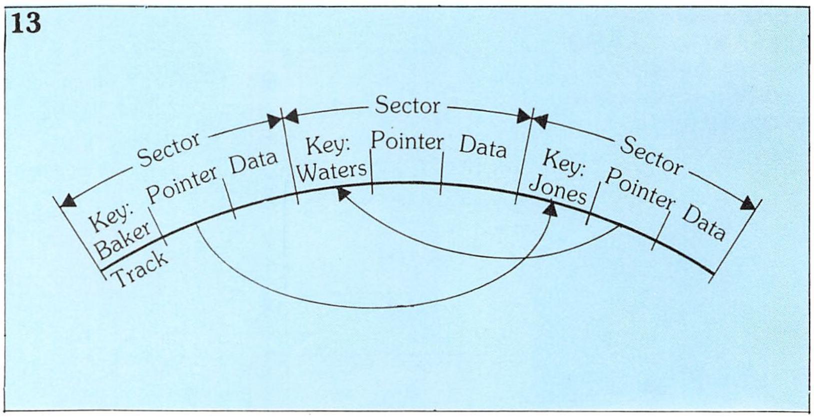
Now, we can look at a practical example of how the linked lists of data items in a sequential disk file can be used to dynamically assign disk space when adding or deleting records. Suppose that the file contains the list of the members of a club, listed alphabetically by surname as in *figure 13*. The record for Mr. Jones can be added easily by giving him a record taken from the space list and updating the pointers to retain alphabetical order. If a member resigns and his name is to be

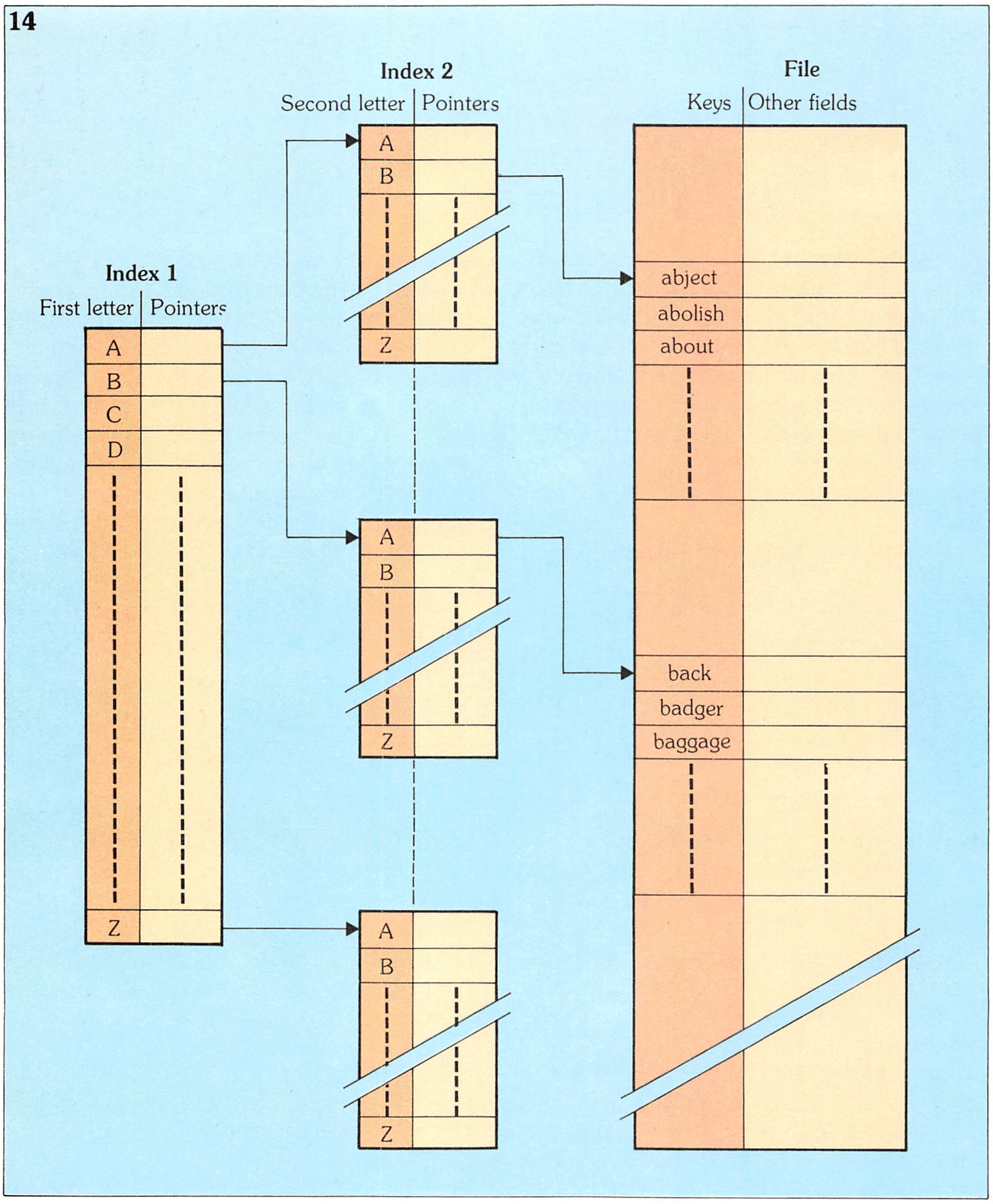
12. Record B is to be deleted from the data block and is automatically returned to the top of the space list.



cancelled, the sector which was assigned to hold his details is returned to the space list and pointers updated.

When a record is added, it is convenient to use a sector on the same track of that file into which it is to be inserted. If another track was used, the disk reading head will have to move thereby increasing access times considerably. To minimise the physical distance between consecutive records, several space lists can be used, each of which is associated with a nearby group of tracks.





13. The record for Mr Jones is added by giving him a record taken from the space list and updating the pointers to retain alphabetical order.

14. Index sequential file where records are alphabetically sequenced according to the keys.

Indexed sequential files

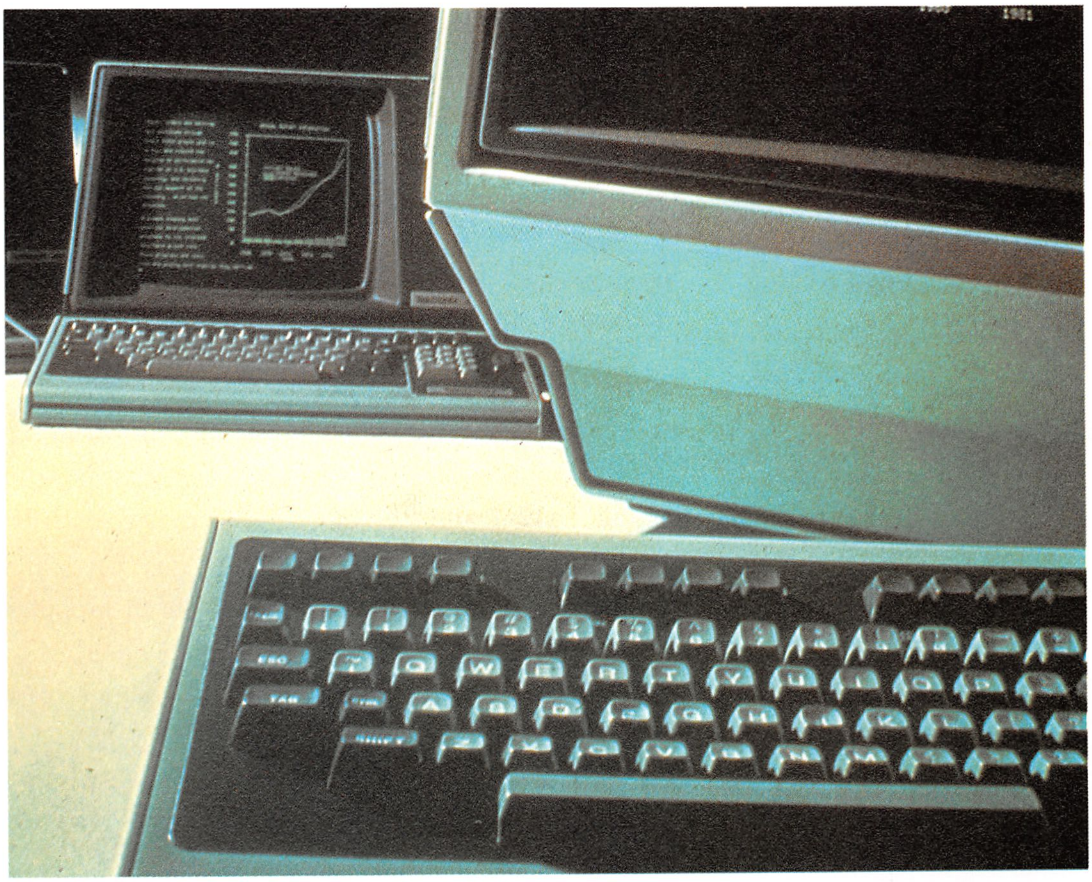
In the **indexed sequential** method of disk storage, records are stored in a sequence identical to that of the ordinary sequential method, but an **index** is automatically written onto the disk. The index contains references to data stored on the disk allowing rapid access to particular records.

So, files can be accessed in one of two ways: either sequentially, with the use of one or more space lists; or in a more rapid way, using the index associated with each file for reference then going to the data record.

The ideal index is obviously one which supplies the *exact* address of the record sought, so that the record can then be read, in a single reading. However, an index containing all the information to do this would take up a great deal of valuable space. To simplify this problem the file can be organised with more than one index, in cascade.

Figure 14 is an example of a system whose records are alphabetically sequenced according to the keys. The first index is simply a list of the 26 letters of the alphabet, each letter referenced to the disk space where keys beginning with that letter are stored. A second index is then similarly used (for the second letter of the key) to find the area of disk in which the keys with chosen first and second letters are stored, and so on.

Index tables, perhaps at several levels, are stored on the disk as ordinary sequential files. When the system is in use the tables are read into central memory and the computer software then uses them automatically when accessing the disk.



Right: example of a terminal being used for financial applications. (Photo: Tektronix).

Randomly organised disk files

In the **random** method there is no direct relationship between logical and physical records. The physical storage address is determined by applying an algorithm to the record key.

Particular records can be accessed at a later date simply by applying the same algorithm and generating its actual address.

You will be able to appreciate that it is

very important to carefully choose file organisational and file handling methods in auxiliary memory equipment. The best way is usually that which allows fastest access to the desired record, with the fewest intermediate readings, and with the least wasted space. Also, an important feature of a good auxiliary store is that files are easily updated, corrected or erased – preferably without the need to rewrite the whole file.

However, faster access times and ease of modification of files are normally only gained with a subsequent higher cost.

Glossary

auxiliary memory	those forms of computer memory which are not part of the computer's central memory
field	a data item within a record. Each field consists of a number of characters
file	organised collection of data records required for providing information in a computer system
inter-block gap (IBG)	a gap of unrecorded tape between blocks of records, to allow for the mechanical inertia of the tape drive system as it stops and starts
inter-record gap (IRG)	a gap of unrecorded tape between records (see inter-block gap)
key field	a specific field in a record, generally the first, which is used to identify the record
labels	recorded information at the beginning and end of a file serving organisational purposes, e.g. identifying the file
linked list	record storage on a magnetic disk in which a pointer indicates the address of the next record
multi-volume file	a file, composed of more than one reel of magnetic tape, or more than one magnetic disk
record	collection of data within a file. Records can be of fixed or variable length

ELECTRICAL TECHNOLOGY

The magnetic circuit

1. Lines of flux flowing through **(a)** an iron core and an air gap; **(b)** the core of a transformer.

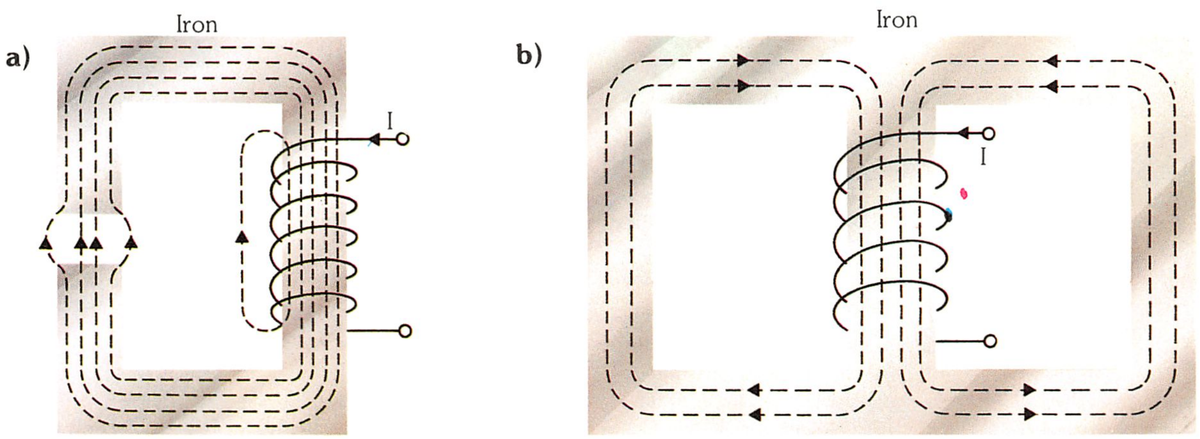
2. Lines of flux flowing through a single type of material in a toroid.

We now know that when a magnetic field acts on ferromagnetic material, the resulting magnetic flux is proportional to the field strength. This is also the case for magnetic systems that do not include ferromagnetics, and it will be useful to calculate the magnitude of the flux in a system which may be made of a variety of materials.

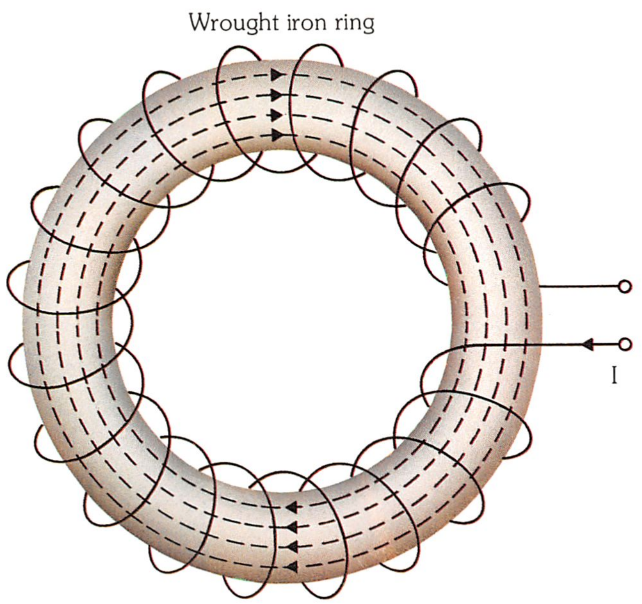
Figure 1a shows a magnetic system where the lines of flux flow through an iron core and an air gap. The permeabilities of air and iron are very different, however, the lines of flux are continuous, therefore the total magnetic flux of the iron must be the same as that of the air gap.

In *figure 1b,* which could represent the core of a transformer, the lines of flux in the central limb of the core, divide and pass around the two side limbs before meeting again. There can be no break in the lines of flux and lines cannot merge, so the sum of the fluxes in the outer limbs must be equal to the total flux in the centre. *Figure 2,* on the other hand, illustrates a simple magnetic system where the flux flows through a single type of material. We need to develop a method of analysis that is applicable to all these different systems.

1



2



Magnetic systems are more correctly called **magnetic circuits** because of their similarity to electrical circuits. In an electric circuit, the electromotive force drives current through various paths, each of which have different properties – indicated by their resistance. In a *magnetic* circuit, the **magnetomotive force** (which is responsible for establishing the magnetizing field strength) drives a **magnetic flux** through a number of paths made up of various magnetic materials. The property of magnetic material that is analogous to electrical resistance is called **reluctance.**

Reluctance

To recap the main points of the chapter dealing with ferromagnetic materials, we know that flux density (B) and field strength (H) are related as follows:

$$B = \mu_r \times \mu_o \times H$$

magnetomotive force (M) and magnetic field strength (H) are related by:

$$M = H \times l$$

and flux density (B) and the total flux, Φ, across an area, A, are related by:

$$\Phi = B \times A$$

Combining these three equations gives:

$$M = \frac{B}{\mu_r \times \mu_o} \times l$$
$$= \frac{l}{\mu_r \times \mu_o \times A} \times \Phi$$

which can be simplified to:

$$M = S \times \Phi$$

where S is the reluctance of the magnetic circuit. From the earlier equation we can see that:

$$S = \frac{l}{\mu_r \times \mu_o \times A}$$

So the reluctance of a magnetic circuit is defined as the ratio of the magnetomotive force to the total flux. Since the magnetomotive force (M) is measured in ampere-turns and flux (Φ) in webers, the unit of reluctance is At Wb^{-1}.

We can now examine some simple magnetic circuits. First of all a simple wrought iron ring, like the one in *figure 2*. This has a mean diameter of 10 cm and a cross-sectional area of 1.5 cm^2. Suppose there are 120 turns of wire around the ring, to find the current that would give a flux density of 0.6 T:

$$\Phi = B \times A$$

and since B = 0.6 T:

$$\Phi = 0.6 \times 1.5 \times 10^{-4}$$
$$= 0.9 \times 10^{-4} \text{ Wb}$$

The mean circumference of the toroid:

$$l = \pi \times 10^{-1} \text{m}$$
$$= 0.314 \text{ m}$$

Using the table of flux and permeability given in the last chapter, we find that at B = 0.6 T, the relative permeability (μ_r) of wrought iron is 2810. The permeability of a vacuum (μ_o) is of course 1.257×10^{-6} Hm.

Using this data we can calculate the reluctance (S) of the magnetic circuit:

$$S = \frac{0.314}{2810 \times 1.257 \times 10^{-6} \times 1.5 \times 10^{-4}}$$
$$= 593 \times 10^3 \text{ At Wb}^{-1}$$

The magnetomotive force:

$$M = 593 \times 10^3 \times 0.9 \times 10^{-4}$$
$$= 53 \text{ At}$$

As M = N × I, we find the current needed to give a flux density of 0.6 T:

$$I = \frac{53}{120}$$
$$= 0.44 \text{ A}$$

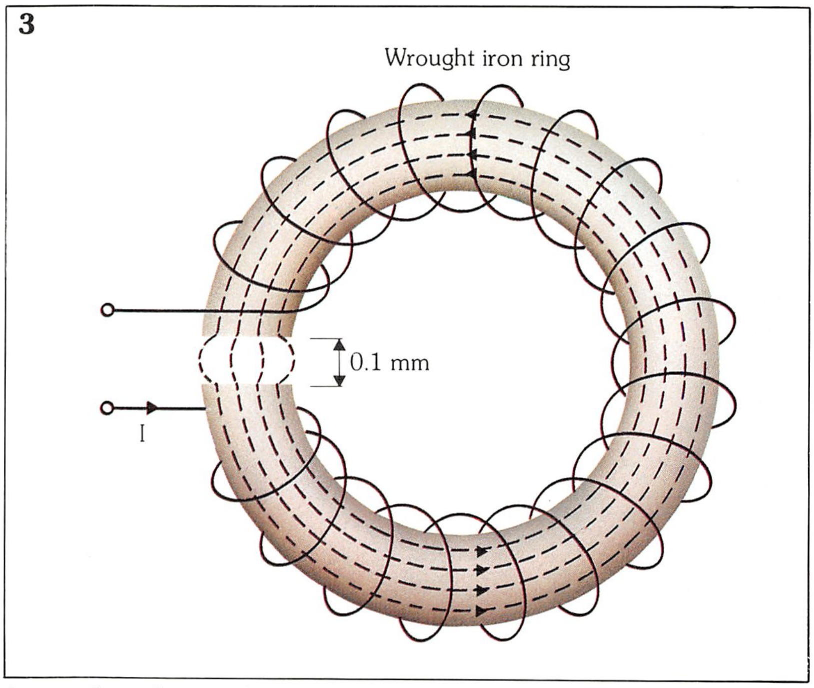
Reluctances in series

If two reluctances, S_1 and S_2, are connected in series, they can be added in the same way as resistances to give a total reluctance S, where:

$$S = S_1 + S_2$$

We can see how this works in practice by looking at the magnetic circuit shown in *figure 3*. Here, the wrought iron toroid in *figure 2* has a gap of 0.1 mm. The dimensions of the toroid, the number of turns of wire and the flux density needed remain the same.

3. A wrought iron toroid with a gap of 0.1 mm.



As the narrow gap will hardly affect the length of the average circumference, the reluctance of the iron (S_{iron}) will still be 593×10^3 At Wb^{-1}. The reluctance of the air gap is given by:

$$S_{air} = \frac{0.1 \times 10^{-3}}{1.257 \times 10^{-6} \times 1.5 \times 10^{-4}}$$
$$= 530 \times 10^3 \text{ At Wb}^{-1}$$

The total reluctance, S, is given by:

$$S = S_{iron} + S_{air}$$
$$= 593 \times 10^3 + 530 \times 10^3$$
$$= 1123 \times 10^3 \text{ At Wb}^{-1}$$

For a flux of 0.9×10^{-4} Wb, the magnetomotive force is given by:

$$M = H \times l$$
$$= 1123 \times 10^3 \times 0.9 \times 10^{-4}$$

Science Photo Library/Jerry Mason

Right: inside the transformer limbs which hold the Torus at JET (Joint European Torus), the EEC's joint undertaking at Culham, U.K. This photograph was taken during construction. When complete, scientists hope to prove that laboratory controlled fusion reactions (like those producing energy in the sun) are possible. Magnetic fields are used to contain and compress plasma (gas heated to very high temperatures) as no material is yet able do so.

$= 101.1$ At

The current required through the 120 turns can therefore be found by:

$$I = \frac{M}{N} = \frac{101.1}{120} = 0.84\ \text{A}$$

Reluctances in parallel

Reluctances can also be combined in parallel, again, in the same way as resistances. The combined reluctance, S of two reluctances, S_1 and S_2, connected in parallel is given by:

$$\frac{1}{S} = \frac{1}{S_1} + \frac{1}{S_2}$$

Consider the magnetic circuit in *figure 1b*, it is operating at a flux density of 0.6 T, with a relative permeability of 2810. The mean length of each of the two outer limbs, including the top and bottom sections, is 25 cm. We find that the reluctance of the left hand limb, S_1, is given by:

$$S_1 = \frac{0.25}{2810 \times 1.25 \times 10^{-6} \times 1.2 \times 10^{-4}} = 590 \times 10^3\ \text{At Wb}^{-1}$$

The reluctance of the right hand limb, S_2, will be exactly the same, so the total reluctance of these two parallel branches can be found:

$$\frac{1}{S} = \frac{1}{590 \times 10^3} + \frac{1}{590 \times 10^3} = \frac{2}{590 \times 10^3}$$

Therefore $S = 295 \times 10^3$. □

DIGITAL ELECTRONICS

Comparing digital and analogue systems

Digital vs analogue

We have seen that any digital system follows a basic pattern, known as the Universal Systems Organisation, in which external forms of information are sensed, a decision is made, and finally an action is taken. The decision may often be made with reference to a memory. *Figure 33* of *Digital Electronics 1* summarised this kind of organisation in a digital system, using a calculator as an example.

Let's now evaluate the nature of the digital approach to building systems, its advantages and disadvantages, by comparing it to the only other approach – analogue.

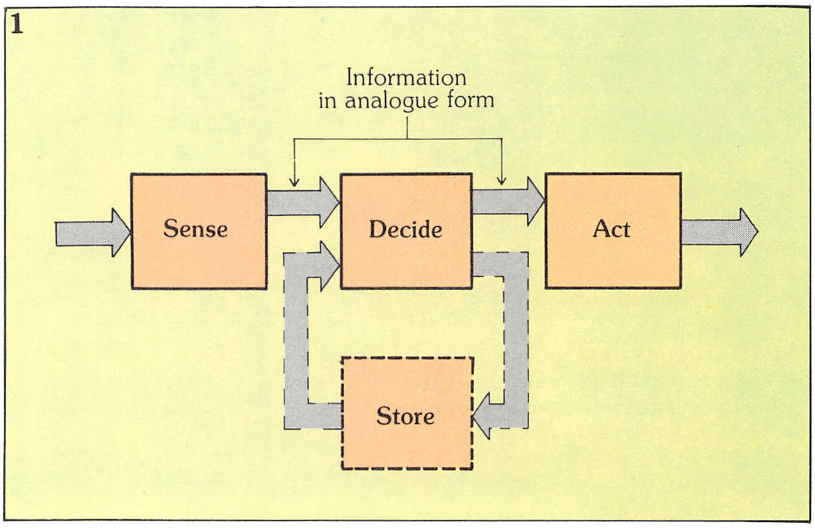
What is an analogue system?
Figure 1 shows the Universal Systems Organisation idea applied to analogue systems. It is similar to that of a digital system with one important difference – the information is in analogue form. (In comparison with digital systems where information is in digital form.)

Digital information is made up of *separate* parts or bits, represented in, say, a computer by different voltage levels which in turn represent only two logic levels, 1 and 0. Analogue information, on the other hand, comprises a *continuous* (or analogous) representation of the original information. Such systems are often termed **linear**.

To represent the analogous or linear information in an electrical or electronic system, some property of electricity, such as voltage or current, must be carefully controlled so as to be an exact copy or model of the original information. An example of an analogue system forming the basis of car petrol gauges or fuel level meters is shown in *figure 2*.

A float on a vertically moving, swinging arm in the petrol tank adjusts a **variable resistor** so that the resistance is directly proportional to the level of petrol: when the tank is full, the variable resistance is low; when the tank is empty, the resistance is high.

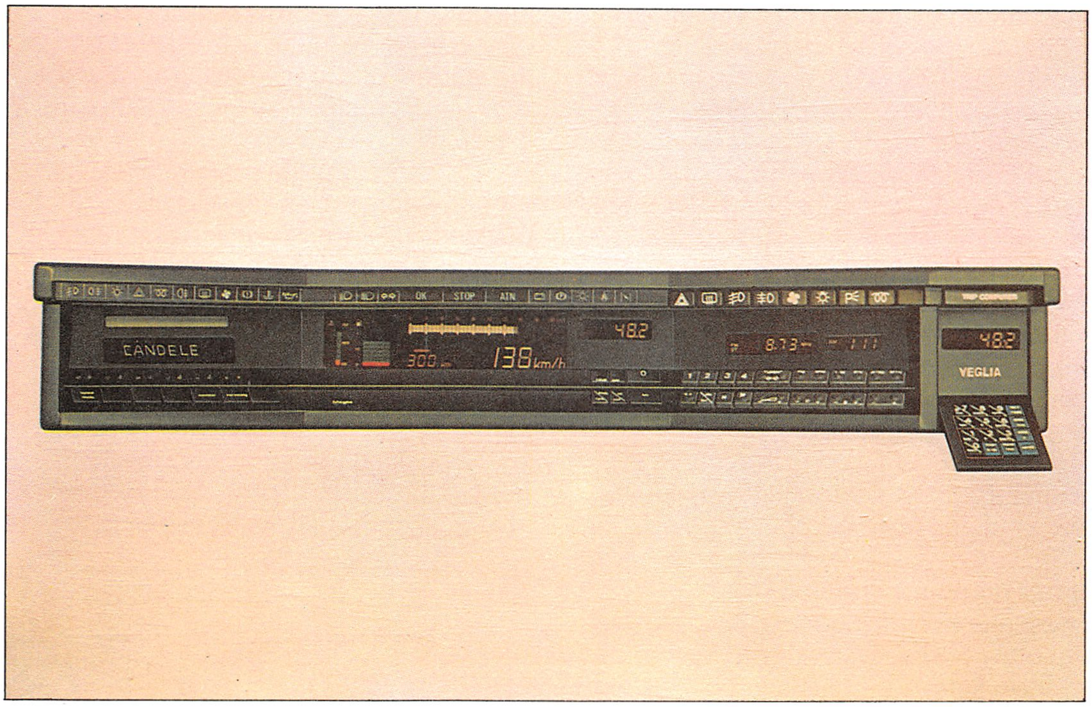
Because the variable resistor is connected to the car battery via the swinging arm, a variable current flows in the wire to the instrument panel (remember Ohm's law: $I = V/R$). For example, 1 mA of current might mean empty and 9 mA might mean full (current levels in between would represent petrol levels from empty to full). At the instrument panel, the pointer needle of the current meter (**ammeter**) indicates petrol level as it moves between the two extremes of measured current.



1. The Universal Systems Organisation idea applied to analogue systems.

Note that the current, in this example, is not switched on and off as in a digital system, but flows all the time the car ignition is on, and varies over the range 1 to 9 mA depending on the petrol level. The pointer in the meter thus copies the movement of the float over its full range. The electric *current* has therefore carried information from one place to another in an *analogue* fashion.

This example can be seen to follow



Right: A digital panel from the dashboard of a car – these are rapidly replacing the classical analogue instruments of rev. counter, speed-ometer, fuel guage etc. (Photo: Veglia Borletti).

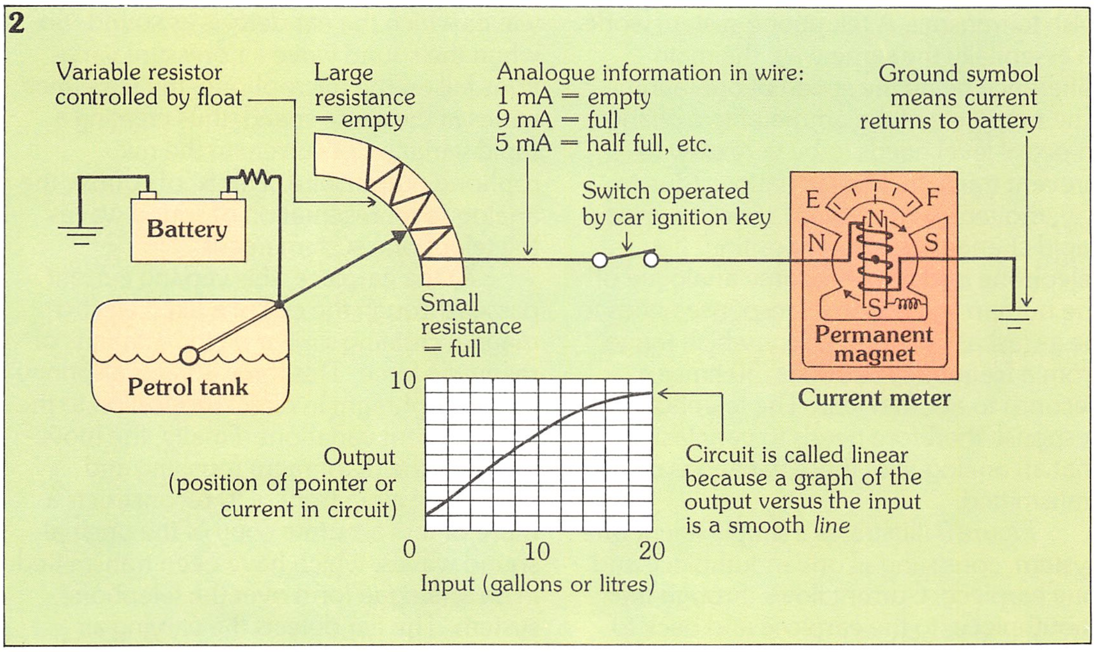
the Universal Systems Organisation approach in that information has been *sensed* (by the float on the swinging arm), *decided upon* (by the variable resistor) and finally *acted upon* (by the ammeter).

The parts of the analogue system

Should you be unfamiliar with the parts of the system described in *figure 2* we ought to describe how they work. The variable resistor would typically be a piece of carbon formed into a circular shape and touched by a moveable contact. Sliding this contact around the carbon shortens or lengthens the path travelled by the electric current.

A second resistor (of a fixed value) between the variable resistor and the car battery **limits** the current in the circuit to within the meter's range (i.e. up to 9 mA).

The **ground** symbols at the battery and meter indicate that these two points



2. An analogue system forming the basis of a petrol gauge. The graph illustrates the special feature of analogue systems – that of varying smoothly over a range.

are connected through a common path – in this case the car's metal chassis – considered to be at 0 V. A complete circuit, needed to allow a flow of current, is thus provided.

Finally, the ammeter acts as a tiny electric motor – the rotor being prevented from turning very far by a spring. The greater the current, the stronger the magnetic field produced in the rotating coil, and the more strongly the N and S poles of the coil and the permanent magnet interact to turn the coil. (Remember, like poles repel, unlike poles attract.) The pointer is attached to the coil, so the greater the current through the meter the further up the scale the pointer points.

Is this analogue circuit linear?

This feature of analogue systems – to vary smoothly over a range (shown by the graph in *figure 2*) – illustrates how the term *linear* is derived. The circuit **output** (the pointer position, or current value in mA) *varies* with the **input** petrol rather than being switched, so the graph is a smooth line, with no sudden jumps. (Don't confuse this with the other meaning of linear, i.e. a straight line.)

A telephone system

The car petrol gauge is a model showing how a controllable property of electricity, current, can be used as an analogue, or direct representation of the information we wish to transmit. A telephone system works in essentially the same way, the main difference being the speed of operation. The **response** of the ammeter to a change in petrol level needs to be very slow to prevent transient level variations (due to, say, movement of the car) showing up as rapid changes in pointer position. If a telephone is to transmit some analogue of the human voice then its response needs to be as fast as vocal variations which range from a **frequency** of 50 Hz (50 times a second) to about 6 kHz. The telephone response therefore needs to very fast so that an analogue of these variations can be transmitted.

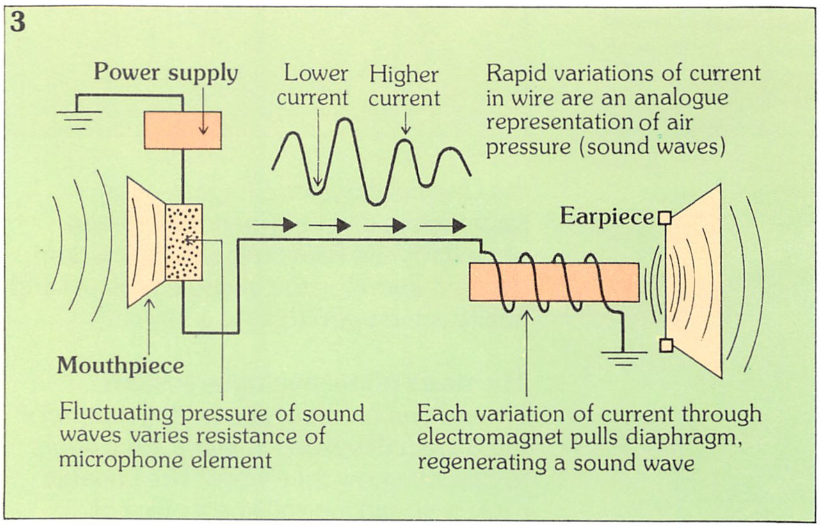
Figure 3 illustrates a simple telephone system, consisting of one mouthpiece and one earpiece. Current flows through the mouthpiece, to the earpiece and back to the power supply through the ground connections. As in the petrol gauge system, the amount of current flowing around the circuit is determined by the resistance in the circuit.

The mouthpiece contains a microphone element – a capsule filled with powdered carbon which acts as a variable resistor. Resistance of the element falls when the powdered carbon is squeezed by air pressure, allowing more current to flow through the circuit. This type of microphone is known as a **carbon granule** microphone.

Speaking varies the air pressure in front of the mouth. These rapid air pressure variations (50 Hz to 6 kHz) form the waves which the ear detects as sound. So when the sound wave air pressure variations fall on the microphone, the resistance varies at the same speed, thus creating a rapid variation of current in the microphone output wire. This is, of course, the analogue representation of sound waves the telephone system needs.

At the earpiece, this varying current passes through the coil of a fixed electromagnet creating similar rapid variations of magnetic force. This force attracts a springy metal diaphragm in direct proportion to the rapid current variations. Finally, the movements of the diaphragm (pushing and pulling the air in front of it) reconstructs a more or less accurate copy of the original sound waves, which have been transmitted in an analogue form over the telephone system. The ear detects the varying air

3. A simple telephone system. Current flows through the mouthpiece, to the earpiece and back to the power supply through the ground connections.



pressure of the sound waves and hears the sound.

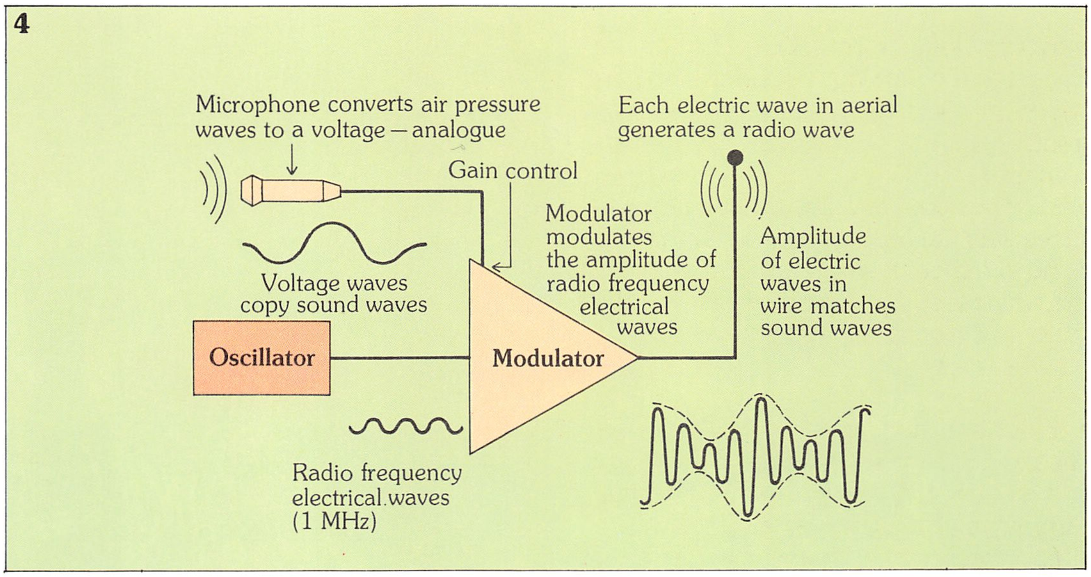
The difference between current and voltage analogues

Although the two examples of analogue systems discussed so far have utilised the control of current as the analogous information method, voltage is often used in other systems. However, the two methods are so similar in concept that the difference is of no real concern. Current is simply the amount or volume of electricity which flows, while voltage is the pressure of electricity – the force that moves the current.

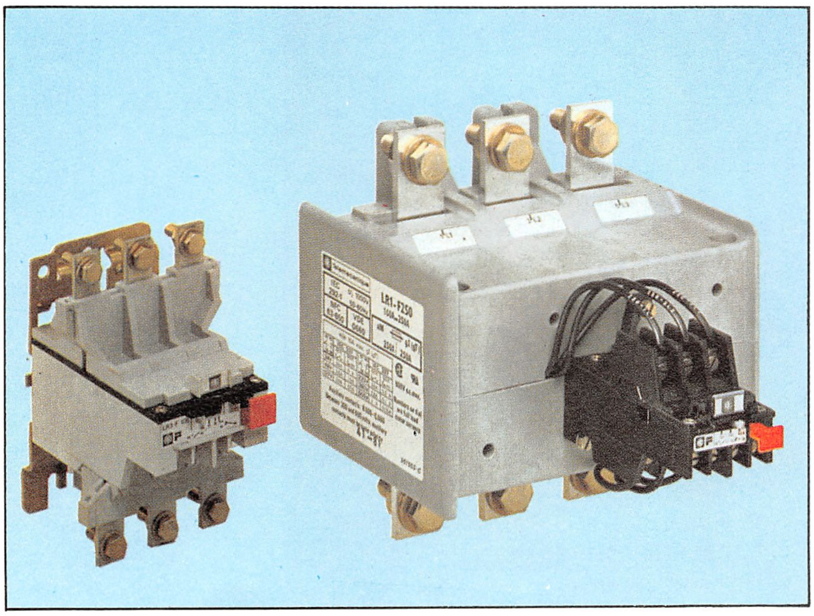
Another analogue system

An analogue system need not only use the simple electrical characteristic of voltage or current as the controlled method of information transmission. It is possible to transmit the information in more complex methods such as the one shown in *figure 4*. The system is an AM radio transmitter, of the kind used in radio broadcasts. AM is an abbreviation of **amplitude modulation**, which is one of the more advanced forms of analogue information.

This technique transmits sound waves (the frequencies of which are low, from say 20 Hz to 15 kHz) by using radio waves of a relatively high frequency – the



4. An AM radio transmitter which transmits sound waves by using high frequency radio waves. The transmitter modulates the amplitude of the radio waves in a pattern which directly represents the sound waves.



Below: an example of tripole relays used for thermal protection. (Photo: Telemecanique).

example radio frequency is 1 MHz. The transmitter simply **modulates**, i.e. varies, the **amplitude**, i.e. height, of the radio waves in a pattern which directly represents the sound waves.

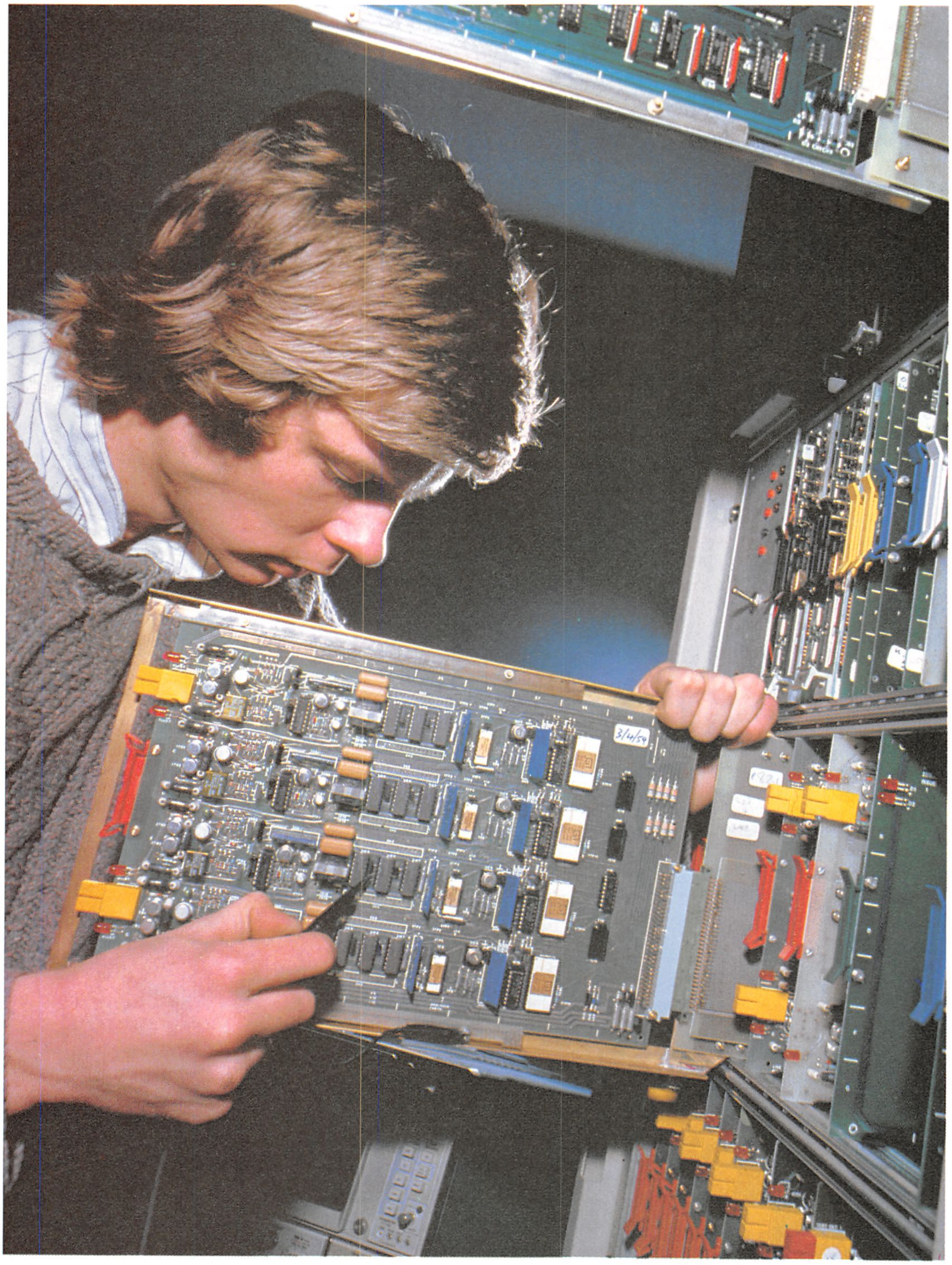
To transmit sound waves by radio the system uses an **oscillator** to generate a 1 MHz electrical wave at a **constant amplitude** – the voltage of the wave goes up and down smoothly a million times a second, the *same amount each time*. A special circuit known as a **modulator** then acts as a variable gain amplifier of these waves. Amplification is a process whereby the input voltage is multiplied by a certain factor, producing taller and stronger voltages at the output. The symbol used, i.e. a triangle, is the standard block symbol used to denote an amplifier.

Now, the gain of the amplifier – the factor by which it multiplies the input – is controlled by the voltage signal from the microphone which is, of course, the voltage analogue of the sound waves striking the microphone. So, the effect of the microphone signal is as though a volume control knob on the amplifier had been turned up and down, thus modulating the amplitude of the 1 MHz output waves.

On reaching the aerial, each electric wave generates a radio wave. The radio waves, which propogate through air, are thus amplitude modulated with the sound waves picked up by the microphone.

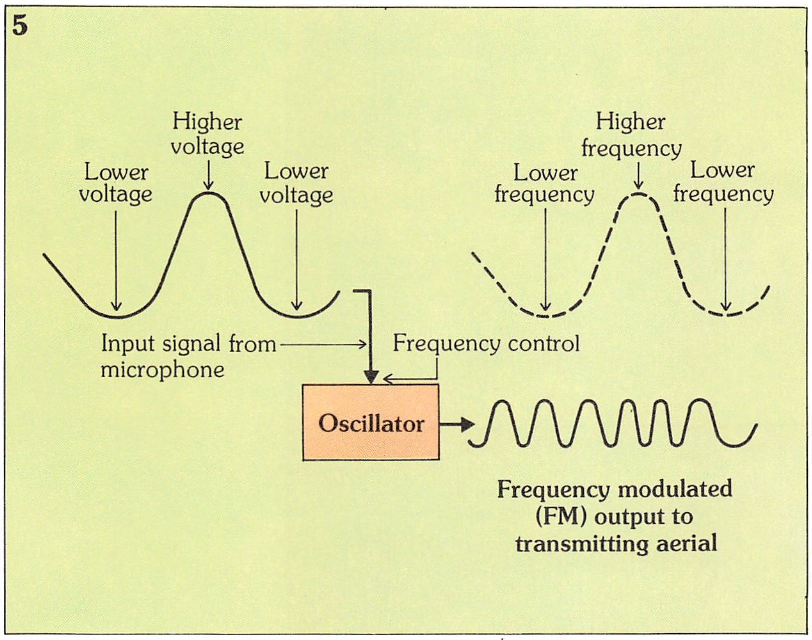
Analogue frequency modulation

Another variation of the analogue techni-



Below: a new digital telephone exchange. These are replacing the old mechanical analogue exchanges.

The Research House/British Telecom



que, similar to amplitude modulation, is **frequency modulation** (FM) which is the basis for certain kinds of high quality radio communications, including TV sound.

As shown in *figure 5*, frequency modulated electric waves are produced by controlling the frequency at which an oscillator varies its output voltage. If these waves are sent to an aerial, corresponding FM radio waves are generated.

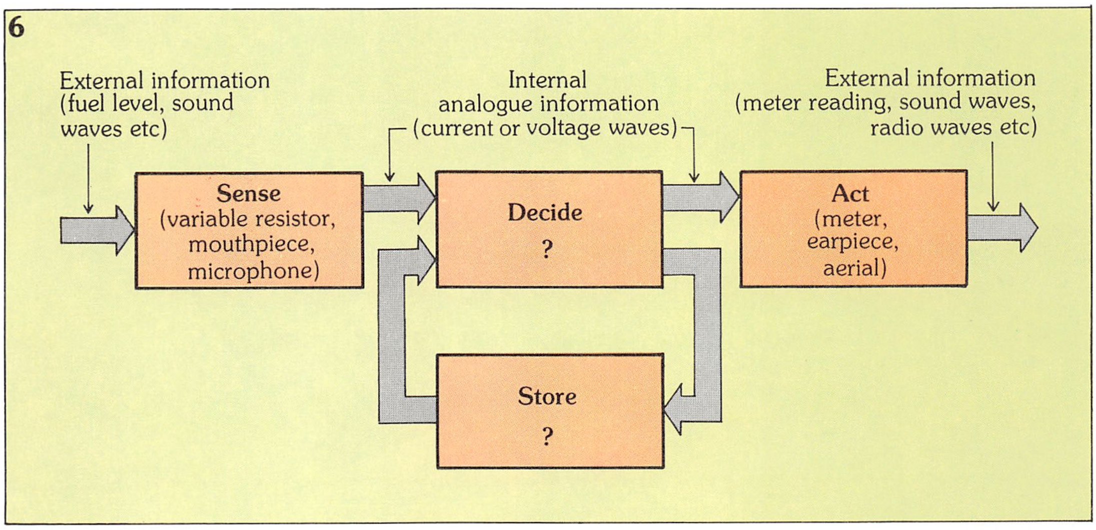
FM signals are used in high quality broadcast systems because they are less prone to interference by noise than AM signals. (This is because noise mainly affects the signal amplitude.) However, FM signals present problems for generation and reception.

Universal systems

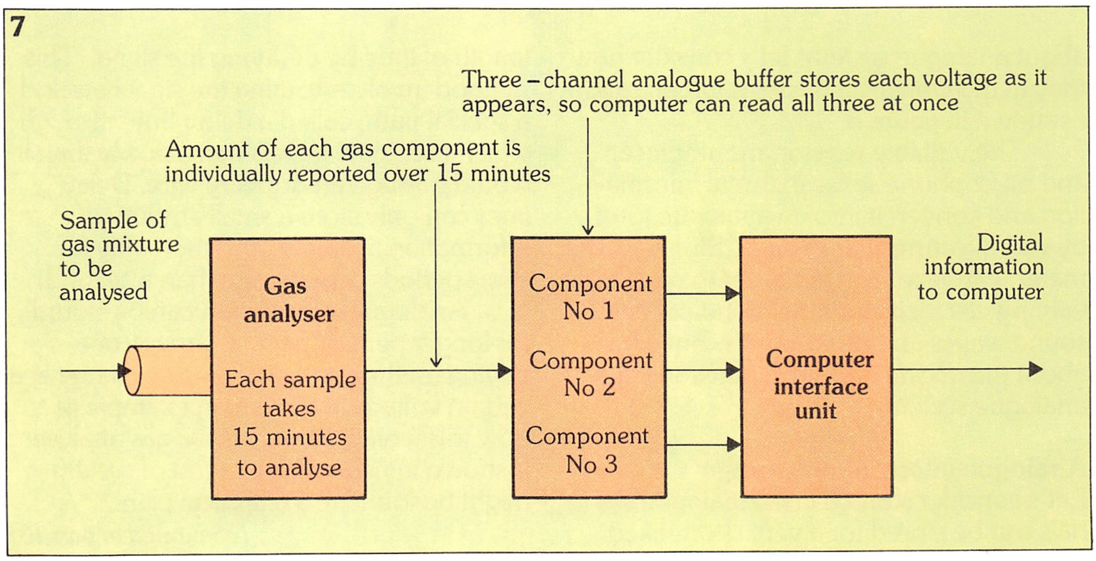
As a reminder of what we've learned so far

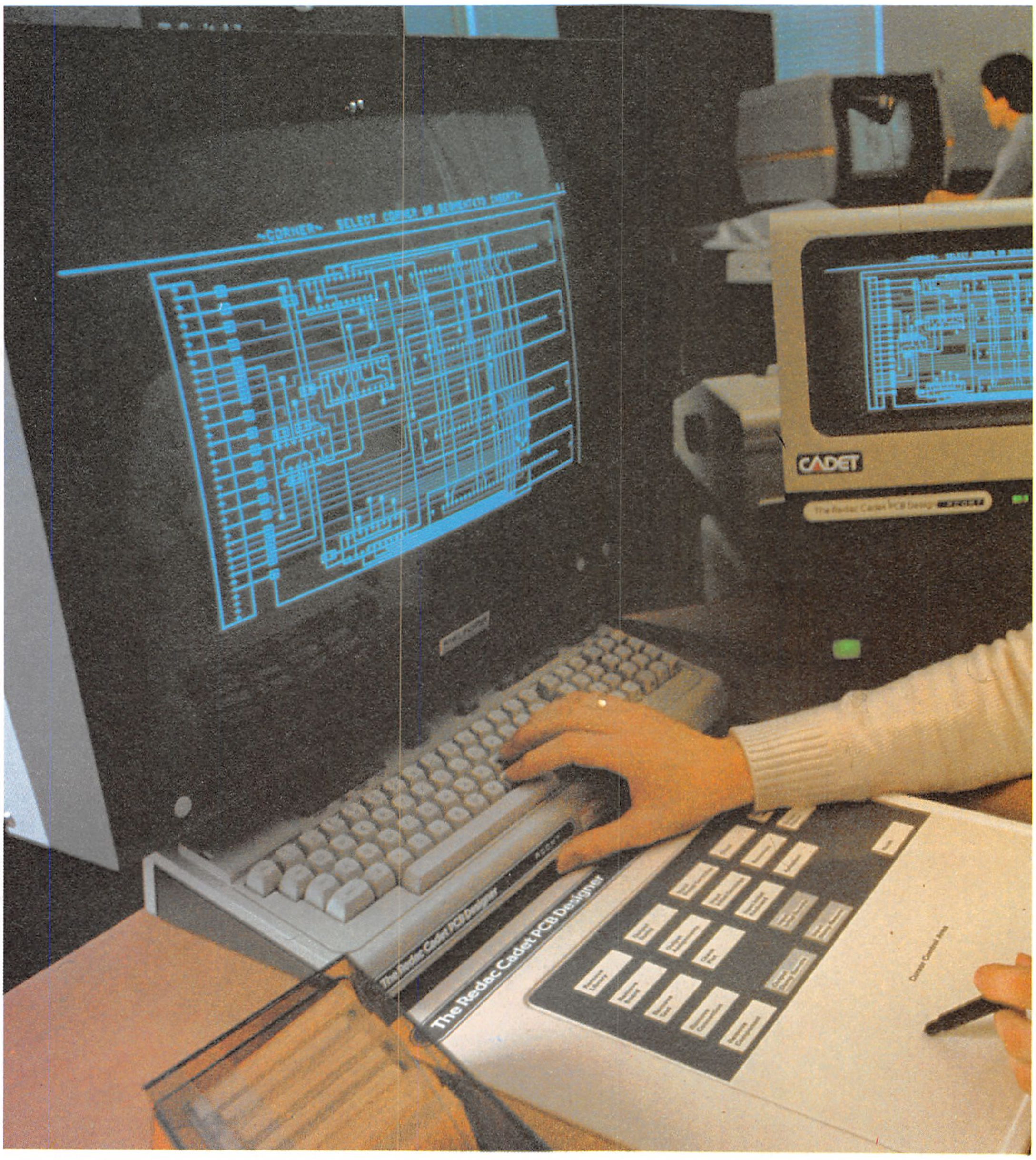
5. Frequency modulated electric waves are produced by controlling the frequency at which an oscillator varies its output voltage. When these waves are sent to an aerial, FM radio waves are generated.

6. How the analogue systems studied so far fit the Universal Systems Organisation.



7. Analogue information can be stored in a capacitor – this method relies on the storage of electric charge.





The Research House/British Telecom

Above: designing a printed circuit board using a graphics tablet and pen.

about analogue systems let's consider how they fit the Universal Systems Organisation as shown in *figure 6*.

The variable resistor, mouthpiece, and microphone *sense* external information and convert it into an analogue form by varying current or voltage. Similarly, the meter, earpiece, and aerial *act* to convert varying electricity into meter indications, sound waves and radio waves. But what about the *decide* and *store* functions in analogue systems?

Analogue information storage

Let's consider storage first. Analogue signals can be stored for a very short, fixed length of time by **delaying** the signal. This method involves routing the signal around a special path, called a **delay line**, through which the signal travels more slowly than it would through an ordinary wire. Delay lines can only store a small amount of information at a time, and then only for short periods (usually less than a second).

Analogue information can be stored for longer periods using a **capacitor** – relying on the storage of electric charge at a certain voltage level. For an example of how this works, an automatic gas analyser is shown in *figure 7*. This kind of system might be found in a chemical plant.

(continued in part 13)